THE UNIVERSITY
of EDINBURGH

Neuropsychological predictors of conversion from amnestic Mild Cognitive Impairment (aMCI) to dementia: a 4-year clinic-based longitudinal study

Jane Lonie

BA (Hons), Msc, Clinical Neuropsychology

Macquarie University Sydney

Thesis presented for the degree of Doctor of Philosophy

College of Medicine and Veterinary Medicine

The University of Edinburgh

2010

Declaration

I declare that this thesis is of my own composition, and that the material contained within describes my own work. I was responsible for the identification of the aMCI subjects, the vast majority of baseline (as well as a number of follow-up) cognitive assessments undertaken by the aMCI patient group, as well as the hypothesis generation, statistical analysis and write-up. Baseline (and in the case of Healthy Elderly Controls, follow-up) assessments for the remaining patient groups (i.e. Alzheimers Disease, Depression and Healthy Elderly Controls) were undertaken by assistant psychologists Claire Donaghue, Lucie Herrman & Mario Parra Rodriguez. Assistant psychologists were also responsible for maintaining the study database (designed and constructed by myself). The thesis has not been submitted for any other degree or professional qualification. All quotations have been distinguished by quotation marks and the sources of information acknowledged.

Jane Lonie, 2010

TABLES

FIGURES

Contents

Papers and Abstracts arising

Lonie JA, Tierney KM, Ebmeier KP. Screening for mild cognitive impairment: a systematic review. *International Journal of Geriatric Psychiatry.* 2009; DOI: 10.1002/gps.2208.

Lonie JA, Herrmann LL, Tierney KM, Donaghey C, O'Carroll RE, Lee A, Ebmeier KP. Lexical and semantic fluency discrepancy scores in aMCI and early Alzheimer's disease. *Journal of Neuropsychology*. 2009;3:79-92.

Lonie JA, Herrmann LL, Donaghey CL, Ebmeier KP. Clinical referral patterns and cognitive profile in mild cognitive impairment. *British Journal of Psychiatry*. 2008 Jan;192:59-64.

Lonie JA, Tierney KM, Herrmann LL, Donaghey C, O'Carroll RE, Lee A, Ebmeier KP. Dual task performance in early Alzheimer's disease, amnestic mild cognitive impairment and depression. *Psychological Medicine*. 2009 Jan;39:23-31.

Lonie, JA, Ukwuori GK, Ebmeier KP. A systematic review of cognitive screening for Mild Cognitive Impairment. *Mind & Brain, the Journal of Psychiatry*. 2009. In press.

Lonie JA, Parra M, Tierney KM, Hermann L, Donaghey C, O'Carroll R, Ebmeier KP. Predicting outcome in mild cognitive impairment. *British Journal of Psychiatry*. 2010. submitted.

Lonie, J.A. Mild Cognitive Impairment: An update. *Proceedings from the Annual Conference for Old Age Psychiatry*. The Radisson Hotel, Glasgow, Scotland. 2006.

Acknowledgements

This work has been supported by the Gordon Small Charitable Trust for research in Old Age Psychiatry. Sincerest thanks go to my supervisors, Professors Klaus Ebmeier, Ronan O'Carroll and Douglas Blackwood for their support and guidance throughout, and to assistant researchers/psychologists Claire Donaghey, Lucie Herrmann, Kevin Tierney and Dr Mario Alfredo Parra Rodriguez for their help with patient recruitment and assessment. I am thankful also to my Royal Edinburgh Hospital clinical colleagues in Old Age Psychiatry, in particular Dr Andrew Lee, for his assistance with patient recruitment, and to Professor Sergio Della Sala, who kindly made available copies of his Dual Task paradigm for inclusion in our neuropsychological research battery.

Abstract

Background

Elderly people who demonstrate memory impairment that falls short of dementia, are referred to as having amnestic Mild Cognitive Impairment (aMCI). AMCI patients have an elevated risk of developing dementia, although not all will do so. Clinical criteria for Alzheimer's Disease (AD) and aMCI do not specify how impairment of a cognitive nature should be defined. The process of differentially diagnosing these conditions can be improved, if knowledge of neuropsychological measures that best discriminate between neurodegenerative and non-neurodegenerative cognitive impairment is used to implement diagnostic criteria for aMCI and AD.

Aims

We sought to 1) determine the frequency of aMCI referrals to our specialist memory clinic, 2) characterise the detailed neuropsychology of a group of patients with aMCI, 3) determine the utility in differential diagnosis and test-retest reliability of these neuropsychological measures, and 4) establish a subset of neuropsychological measures that were of prognostic utility in aMCI.

Methods

The case notes of 187 consecutive referrals received by our Royal Edinburgh Hospital memory assessment service across an 18-month period were reviewed retrospectively and numbers of patients fulfilling aMCI criteria were recorded. The baseline neuropsychological performances of 46 patients with aMCI, 20 patients with very early stage AD, 20 elderly patients with depressive symptoms and 24 healthy elderly participants were compared in order to determine their usefulness in differential diagnosis. AMCI participants were followed-up across an average of 4 years. Baseline neuropsychological performances of the aMCI dementia converters and aMCI non-converters were compared. Logistic regression analysis was applied to ascertain the predictive accuracy of a combination of these.

Results

One quarter of referrals received by our memory assessment service met criteria for aMCI, most of whom displayed additional neuropsychological impairments of a non-memory nature, all the while performing above the highest cut off points on even the most comprehensive dementia screening measures. A number of neuropsychological measures were highly sensitive and specific to early AD however, similar combinations of both high sensitivity and specificity to aMCI were not achieved. Forty one percent of patients presenting to our service who fulfilled criteria for aMCI, received a clinical diagnosis of dementia across an average 4-year period. Performances on a comprehensive cognitive screening measure and a measure of delayed word recognition accuracy at baseline, classified 74% of aMCI patients comprising our clinic sample in accordance with their prognostic fate.

Conclusion

A significant proportion of patients presenting to specialist memory clinics display episodic and semantic memory/ or executive impairment that falls short of dementia and that is not detectable using traditional bedside screening measures. The vast majority of such patients (i.e. 72%) experience persisting or progressive cognitive impairment, and a significant proportion (41%) go on to receive a clinical diagnosis of dementia. The baseline neuropsychological performance of aMCI patients who do and do not develop dementia differs, and contributes over and above clinical information to the prediction of long-term diagnostic outcome. The high frequency with which aMCI is encountered in clinical practice, coupled with the minority of aMCI patients who experience resolution of their cognitive impairment, and the sensitivity and prognostic utility of several neuropsychological tasks, has implications for the clinical management of patients with aMCI.

1. Introduction

Dementia is a syndrome of acquired intellectual impairment of sufficient severity to interfere with social or occupational functioning caused by brain dysfunction (Salmon and Bondi 2009). Current diagnostic criteria for dementia require that a deficit and decline in memory must be present, along with at least one other area of cognitive disturbance (American Psychiatric Association 1994;McKhann et al. 1984), although the cognitive aspects of clinical criteria proposed for the less commonly encountered focal dementia subtypes vary (Mendez et al. 2009;Mesulam 2001;Neary et al. 1998).

The history of dementia can be traced back to just over one hundred years ago to clinical descriptions of a 51-year-old lady with 'pre-senile' dementia in whom the histopathological hallmarks of AD were later discovered. The identification of a disease substrate forced a gradual shift in the conceptualisation of dementia away from that of an unexplained disturbed behavioural state towards a medical syndrome. The later (1976) discovery of identical histopathological changes in the brains of patients with the more commonly encountered 'senile' dementia provided the impetus for large scale research into the cause, neuropathologic features and clinical characteristics of dementia.

In 2007 there were estimated to be 683,597 people with dementia in the UK, equating to one in every 88 persons of the entire UK population (Knapp and Prince 2007). The number of people with dementia within the UK is forecast to increase by 38% over the following 15 years and by 154% over the next 45 years, such that it is expected that there will be 1,735,087 people with dementia residing within the UK by 2051. The prevalence of dementia is known to double every five years up until the age of 84 years, with the largest proportion of sufferers aged between 75-89 years. The present nation-wide total cost of dementia (including formal care agencies as well as the financial value of unpaid informal i.e. family care) is estimated at £17.03 billion. Future costs will obviously rise considerably in line with the projected increase in dementia prevalence.

Reaching a diagnosis of dementia at an early stage is important for a number of reasons. It allows time for the patient and his/her family to plan for the future and to attend, in a pre-emptive manner, to financial and treatment matters. The extent to which the patient can

effectively contribute to such planning will be dependent upon his/her cognitive capabilities (as dictated to a large extent by the stage of the illness). Diagnosing dementia at an early stage provides a context from which the patient and family may understand disease related cognitive and behaviour changes and their functional impact(s), as well as the opportunity for specialists to educate and advise on symptom management. Early diagnosis has also been shown to facilitate caregiver participation and ease caregiver stress, delaying the point at which the patient requires full-time institutionalised care (Weimer and Sager 2009). Pharmacological treatment may also help to slow the progression of symptoms (Lopez-Pousa et al. 2005).

Dementia therefore poses a considerable and increasing challenge to society. Diagnosing the condition at an early stage is both clinically and economically advantageous and likely to become increasingly important with the advent of disease modifying or halting agents.

The clinical application of neuropsychology is concerned with the behavioural expression of brain dysfunction (Lezak 2004). Neuropsychology is applied within clinical settings to assist with identification, differential diagnosis, management and rehabilitation of brain injuries, including brain injury arising within the context of neurodegenerative illness. Whilst viewed as giving rise to 'diffuse' cognitive impairment, neurodegenerative diseases were of limited interest to neuropsychologists who sought to unravel the brain's basis for cognitive function through the use of dissociation techniques. During the early 1980's however, the development of standardised cognitive assessment techniques and clinical criteria for dementia allowed reliable diagnoses to be made at a much earlier disease stage. As a result, traditional conceptualisations of dementia as giving rise to 'generalised impairment of intellect' were replaced with the notion of a syndrome that gave rise to 'symptoms that were reflective of the underlying disease topography'. The neuropsychological study of neurodegenerative disease grew out of this more 'contained' view of neurodegenerative-based brain damage.

Major developments in our understanding of the order of progression of cognitive symptoms in dementia have followed. Neuropsychological research has facilitated the differentiation of cognitive changes that occur as a normal consequence of aging, from those that signal the onset of a dementia syndrome, the differential diagnosis of a wide variety of dementia

subtypes, and the identification and detailed characterisation of the cognitive changes and underlying cognitive processes that comprise the manifesting signs of many of these.

In the case of AD (the most common form of dementia accounting for 62% of all UK dementia diagnoses (Knapp and Prince 2007)), a wealth of neuropsychological evidence shows that episodic memory impairment, characterised by rapid forgetting (Attix & Welsh-Bohmer, 2006) is usually the earliest and most salient aspect of the syndrome and that numerous features of memory task performance (i.e. savings or delayed recall, reduction of retrieval demands, serial position effect, semantic encoding benefits & intrusion errors) can be used to differentiate AD from normal aging with very high degrees of accuracy (Salmon and Bondi 2009). The rapid forgetting that characterises the earliest stages of the AD disease process is most easily demonstrated using neuropsychological tasks that that impose delay intervals, requiring the patient to retain newly learned material across time. By contrast, immediate (as assessed by measures of forward digit span) and long-term memories remain intact in the early stages.

As the neuropathology of AD spreads beyond medial temporal lobe structures into areas of the association cortices of the temporal lobe and frontal lobes (Braak and Braak 1991), aspects of semantic memory, language and attention/executive function become increasingly impaired, making it difficult for patients' to retrieve words and names (of people and places) and to generate these at pace, in accordance with a specified category. It becomes progressively difficult for patients to understand complex or lengthy syntax and grammer, affecting their ability to comprehend spoken language and instruction.

Aspects of frontal or executive functioning, specifically the ability to divide one's attention between two things at pace, to forward plan and problem solve, also become impaired, leading to an increasing loss of functional independence. The emergence of impairments of a visuospatial nature (in the non-aytipical i.e. Posterior Cortical Atrophy or visual variant of AD) is generally thought to arise secondarily to the above changes (Perry and Hodges 2000) and deficits of this nature can be elicited via the administration of complex 2-D copying or self-drawing (i.e. clock drawing) tasks. Impairments of this nature frequently give rise to difficulties with topographical orientation and spatial layout making it increasingly difficult for the patient to venture outwith his/her immediate and familiar environment alone.

Advances in our understanding of the staging of cognitive deficits in AD have formed the platform for the past decade's focus on its' pre-clinical stage. The drive to facilitate AD diagnosis at an earlier stage of the disease course has been fuelled by a number of factors, including evidence in support of the existence of a long prodromal phase to the neurodegenerative illnesses, the availability of drug therapy for AD, and patient requests for prognosis and treatment.

Current clinical criteria for AD require cognitive decline and impairment to be present in at least one domain in addition to memory (American Psychiatric Association 1994;McKhann et al. 1984). Functional impairment is a further requirement. It is apparent however, that impairment(s) in different cognitive domains and functional decline(s) do not emerge simultaneously. It has therefore been argued that the sequential nature of symptom emergence in the earliest AD disease stages is more commensurate with a continuum of cognitive impairment as opposed to the dichotomous (i.e. impairment present or absent) approach enlisted by current diagnostic criteria (Hachinski 2008).

Memory complaints can predate other cognitive symptoms and functional decline by considerable intervals (Amieva et al. 2005). This observation begs the question of whether any aspect(s) of the cognitive performances of persons with memory, but not additional functional impairment, could be used to reliably pre-empt a future diagnosis of AD, or 'reduce the threshold at which we are willing to arrive at a diagnosis of AD in accordance with existing clinical criteria' (Chertkow et al. 2007).

One factor complicating the detection of AD at an earlier stage, is the high prevelance (i.e. 25-50%) of comorbid depression and mild cognitive impairment among the elderly. 'Depression is a mood disorder that produces sadness, negative self-regard, loss of interest in life and disruptions of sleep, apetite, thinking and energy levels' (Steffens and Potter 2008). A depressive episode is diagnosed if, and when, the above symptoms are deemed troublesome enough to be interfering with daily life and they have persisted for more than two weeks.

A majority of studies (Kohler et al. 2009;Rapp et al. 2005;Thomas et al. 2009) although not all (Fischer et al. 2008) report impaired neuropsychological performance among elderly depressed patients relative to age and IQ matched controls.

Cognitive deficits are most commonly noted in the domains of processing speed, episodic memory and effortful tasks tapping aspects of executive functioning (Steffens and Potter 2008), although there is considerable variability among studies, some of which may be attributable to the heterogeneity of cognitive measures employed. Some studies have reported evidence of fairly generaised cognitive deficits across a wide range of domains i.e. executive functioning, processing speed, episodic and semantic memory (Herrmann et al. 2007;Kohler et al. 2009) whilst others have reported more isolated and distinctive profiles of cognitive impairment (Rapp et al. 2005;Thomas et al. 2009) that correspond with age of onset of the first depressive episode.

With specific reference to the latter observation, there is evidence to suggest that late onset depression (LOD) (i.e. when the first depressive episode occurs after the age of 50 yrs) is associated with greater levels of imparment of attention and executive function (Herrmann et al. 2007;Kohler et al. 2009;Rapp et al. 2005) and processing speed (Herrmann et al. 2007) than recurrent or early onset (i.e. first episodic before the age of 50 yrs) depression. Greater levels of executive and episodic memory compromise have been demonstrated in association with elderly depressed patients with an increased number of deep white-matter hyperintensities (Steffens and Potter 2008), and it is hypothesized that the executive deficits in the LOD group may therefore arise, together with depressed mood, as a result of subcortical vascular pathology.

Conversely, some research suggests that memory deficits may be more focally affected among older individuals with a history of recurrent depression beginning earlier in life i.e. EOD (Rapp et al. 2005).It has been argued that such observations are in keeping with findings of significant hippocampal volume loss in EOD which is postulated to arise as a consequence of depression related stress related hypercortisolemia having a toxic effect on the hippocampus, disrupting regulation of glucocorticoid secretion resulting in hippocampal atrophy. Of note, however, is that fact that the episodic memory impairments where documented in sufficient detail, appear to reflect difficulty at the levels of learning and unsupported delayed free recall as opposed to cued recall or recognition (Rapp et al.

2005;Thomas et al. 2009). In addition, there remains uncertainty also as to what extent the cognitive deficits seen in association with depression among the elderly might be underpinned by a more generalized slowing of information processing ability, as has been argued in the case of normal aging (Salthouse 1996).

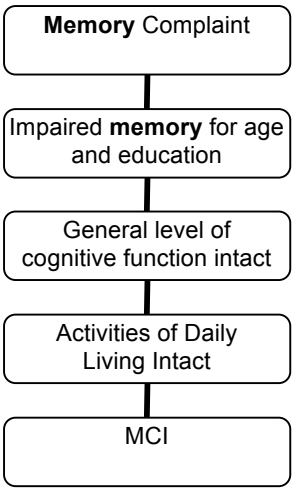
Regardless of the lack of clarity surrounding the neuropsychological profile of deficits accompanying depression in the elderly, cognitive impairments of a significant magnitude are frequently observed among this patient group, the assessement of which consequently forms an integral part of the early and differential diagnosis of AD and aMCI.

Measurement of cognitive function represents just one, all be it an important, approach to detecting and diagnosing AD at a very early and pre-clinical stage. Other work has looked at the ability of imaging (for the most part MRI scanning), biomarkers (i.e. total tau, AB42 & Phospho-tau) and changes of a behavioural nature to predict the future onset of clinically diagnosable AD. A recent meta-analysis of imaging and biomarkers for AD (Schmand et al. 2010) indicated some promise for the CSF markers in so far as there overall predictive accuracy levels were similar to that of memory impairment four years prior to the point of diagnosis. Furthermore, the effect sizes for the CSF markers were largest when assessed longer before the point of diagnosis. However, atrophy of the hippocampus or other MTL structures was found to be a less accurate predictor of future AD than memory impairment, and the largest effect sizes, (which are themselves likely to represent an underestimation owing to the removal of variability inherent in the inclusion of memory impairment as a selection criteria for a majority of studies), were seen association with measures of delayed memory recall. Other forms of neuroimaging i.e. PET and fMRI, were not evaluated as part of this analysis and their comparative contribution to the early and pre-clinical detection of AD therefore remains to be determined. There is also evidence to suggest that behavioural symptoms such as depression (Steffens and Potter 2008) and apathy (Bartolini et al. 2005) predate the onset of AD and may be helpful in detecting the disease in it's pre-clinical stages.

1.1 MCI : Application of criteria and prevalence in specialist memory clinics

The term Mild Cognitive Impairment (MCI) was adopted by Petersen and colleagues (Petersen et al. 1999) who tracked the longitudinal course of a group of 76 cognitively impaired, non-demented patients. After 4 years, 48% had developed dementia, with an annual mean conversion rate of 12% as compared to 1-2% in an age matched community control group. Petersen (Petersen et al. 1999) proposed that these mildly cognitively impaired patients were in a transitional state, between normal ageing and very early Alzheimer's disease (AD). A set of criteria for MCI was established requiring that a patient have (1) a memory complaint, (2) retain normal activities of daily living and (3) normal levels of general cognitive function and (4) display objective evidence of abnormal memory for age (5) fulfilling criteria for dementia.

Figure 1.1 Original Criteria for MCI

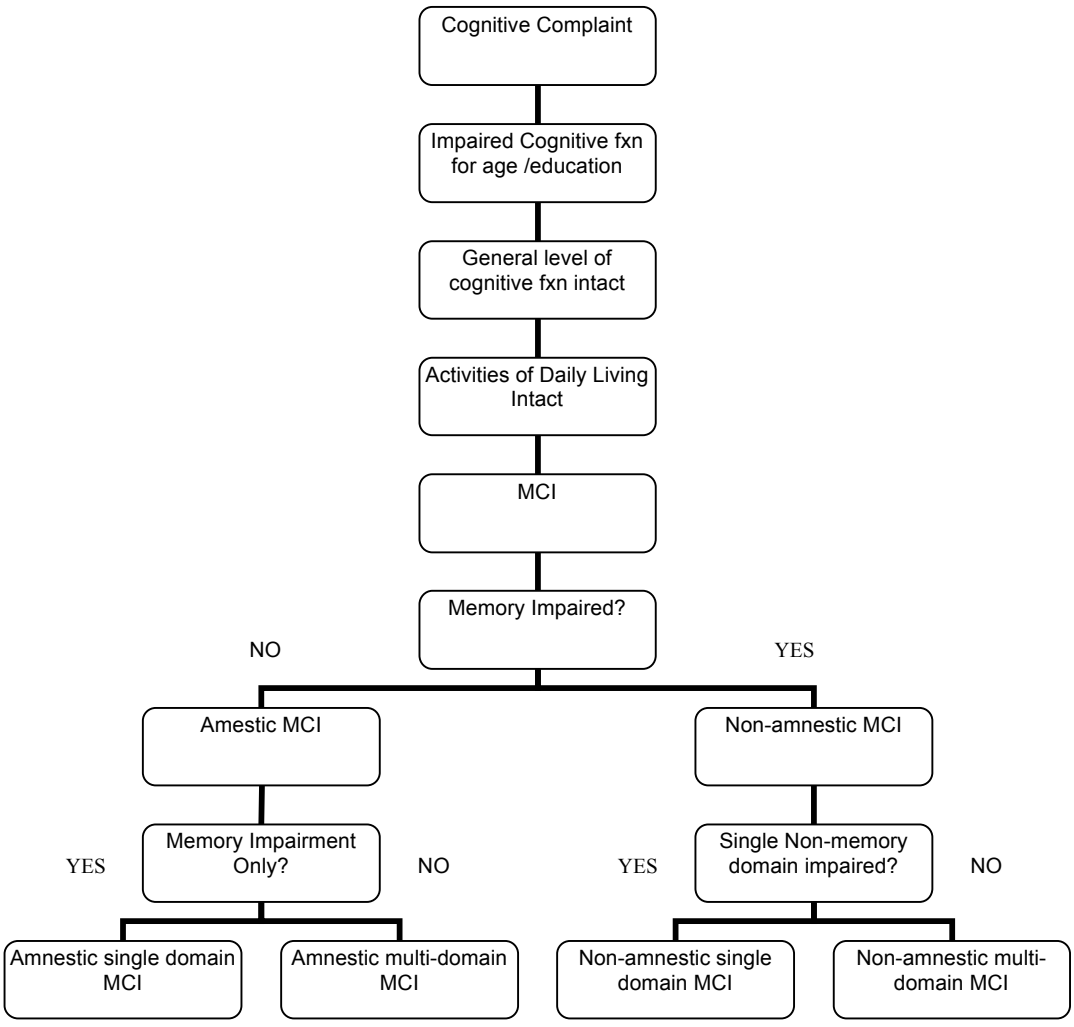


Prior to this, a variety of terms had been used to denote persons with cognitive impairment falling short of dementia (Collie and Maruff 2002). These terms differ from the Petersen MCI criteria on a number of levels including (1) the normative cut off points and (2) comparison groups that are used to define cognitive impairment, (3) the number and type of cognitive domain(s) affected and most importantly (4) the presumed underlying cause(s) of

cognitive impairment. Criteria for mild cognitive impairment (MCI) are unique in that they represent an initial attempt to define features of the dementias in their preclinical phases.

Work subsequent to the publication of initial criteria for MCI indicated that not all elderly persons with cognitive impairment falling short of dementia exhibit an isolated memory complaint (Economou et al. 2007;Gualtieri and Johnson 2005;Loewenstein et al. 2006a;Riberio et al. 2006). The original criteria were expanded to accommodate this (see Figure 1.2).

Figure 1.2 MCI Subtypes



Petersen and colleagues (Petersen 2004;Petersen 2005c) identified four MCI subtypes in accordance with the presence of memory (amnestic) or other forms (non-amnestic) of cognitive deficit, and single or multiple domains of cognitive impairment. It was anticipated that this more precise characterisation of the cognitive presentation of MCI sufferers would facilitate greater comparability of research findings and determination of differing prognoses for the MCI subtypes.

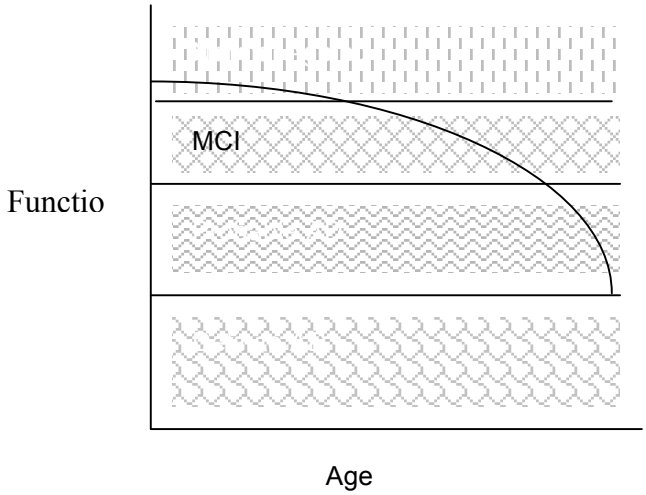
Two further amendments to Petersen's (Petersen et al. 1999) MCI criteria have been proposed subsequent to their initial publication, in recognition of growing evidence to suggest that relatively few MCI sufferers exhibit isolated memory impairment and retain preserved complex activities of daily living. The International Working Group on Mild Cognitive Impairment (Winblad et al. 2004) has recommended that these aspects of the MCI criteria be 'loosened' to allow for some degree of impairment in non-memory and functional performance. One problematic consequence of the recommended changes is a resultant blurring of boundaries between MCI and early AD, as the absence of functional and multi-domain cognitive impairment in MCI could no longer be used to differentiate these patient groups.

Despite the evolving nature of it's criteria, in 2001 MCI was endorsed as a useful clinical construct by the American Academy of Neurology (Petersen 2004) and clinicians were urged to identify and monitor such patients for their increased risk of developing a subsequent dementia. Petersen's original and revised MCI criteria have subsequently gained widespread acceptance among specialist research centres (Petersen and O'Brien 2006) and clinicians (Mitchell et al. 2008) although research has continued to focus on the amnestic single and amnestic multi-domain subtypes. A case for the inclusion of the amnestic MCI subtype in the Diagnostic and Statistical Manual Version 5 (DSM-V) has been made (Petersen and O'Brien 2006) on grounds that there is an increasing need for clinicians to be able to recognize and treat dementias at an earlier stage, and as a clinical entity, MCI fulfils comparable numbers of the criteria that are typically used to classify DSM disorders (Kendell 1989).

Consideration of MCI as a diagnostic entity reinforces the need for clinicians and researchers to be able to differentiate this group of patients from their age contemporaries and from

patients with early AD. Even without diagnostic status, initial findings indicate that between 20% (Lehrner et al. 2005) and 23% (Alladi et al. 2006) of all referrals received at specialist memory clinics comprise patients who fulfil Petersen's MCI criteria. This would suggest that the clinical assessment and wider management of MCI already comprises a significant part of the specialist practitioner's workload, although the extent to which these estimates generalise to other memory clinics remains unclear.

Figure 1.3 Theoretical progression of the transition from normal ageing to Alzheimer's Disease



Probable AD in accordance with NINCDS-ADRDA (McKhann et al. 1984) and DSM-IV (American Psychiatric Association 1994) criteria; Definite AD in accordance with autopsy findings (Mirra et al. 2009); MCI in accordance with Petersen's (Petersen et al. 1999) criterion.

Whilst there is general consensus regarding the plausibility of a long prodromal phase to the dementias, and hence the theoretical construct that underlies MCI (see Figure 1.3), a good deal of controversy persists regarding how best to apply MCI criteria. The bulk of ongoing debate centres around which cognitive measure(s) at which cut off point(s) should be used to objectively establish impairment in memory and other cognitive domains. It is well established that cognitive test sensitivities and specificities to aMCI vary (Alladi et al. 2006;Dierckx and Engelborghs 2007;Grundman et al. 2004;Loewenstein et al. 2006a;Lonie et al. 2008). As a consequence, choice of cognitive measure, or indeed the number of measures (Brooks et al. 2007;De Jager and Budge 2005;Loewenstein et al. 2006b) used to define MCI could conceivably affect both the cognitive and affective make-up of the MCI group, and in turn, longitudinal outcome.

As with test selection, there is relatively little information to inform the selection of cut off values in establishing episodic and non-amnestic cognitive deficits in MCI. With the exception of Grundman's proposed research criteria (Grundman et al. 2004), which arbitrarily specifies a performance of 1.5 SD below age and education matched controls on the delayed recall component of the logical memory subtest from the now outdated Revised Wechsler Memory Scale (WMS-R; (Wechsler 1987)) this issue has not been resolved. The research criteria have not been adopted uniformly, indeed there is evidence to suggest that defining memory impairment in this manner can lead to false positive MCI diagnoses (Brooks et al. 2007). Performance levels of between -1 and -2SD below age norms have typically been applied as a means of establishing episodic memory impairment in MCI. Comparison groups have comprised age and IQ (Alladi et al. 2006;Rentz et al. 2004), age and education matched healthy elderly (Artero et al. 2006;Loewenstein et al. 2007a;Visser and Verhey 2008) or published age matched normative data (Fox et al. 1998;Godbolt et al. 2004).

Elevated rates of conversion from MCI to dementia (relative to general population estimates of 1-2% / annum), have however been reported in association with the use of -1SD (Geslani et al. 2005;Schmidtke and Hermeneit 2007), -1.5SD (Artero et al. 2003), -2SD (Palmer et al. 2007) cut off values to define episodic memory impairment in MCI. There is some evidence to suggest that greater levels of baseline episodic memory impairment are associated with a greater likelihood of progression to dementia during follow-up (Daly et al. 2000;Palmer et al. 2003).

For reasons outlined above, the neuropsychological aspects of the MCI classification have been deemed the most poorly defined (Portet et al. 2006). Some authors have downplayed such concerns by emphasizing that the diagnosis of MCI is 'not a neuropsychological one but rather a judgment call on the part of the clinician' (Petersen et al. 1999;Petersen and O'Brien 2006). It is nonetheless clear that cognitive testing forms a central component of the MCI criteria & hence the wider diagnostic decision-making process. Without specification of precise measures, cut off values and normative comparison groups, variability in the case definition of MCI persists, with the potential to affect outcome and comparability of study findings. The neuropsychological aspects of MCI require further study prior to the establishment of criteria that can be applied clinically in a consistent manner.

On the basis of findings to date, it was predicted that 1) patients who fulfil Petersen's MCI criteria would comprise a significant number of referrals to our Older Adult Neuropsychological Assessment Service and that 2) the bulk of MCI patients would demonstrate memory as well as other domains of cognitive impairment on comprehensive neuropsychological assessment. We further hypothesized that 3) the classification of MCI subjects as normal, single or multi-domain, amnestic or non-amnestic would vary widely in accordance with the neuropsychological measure(s) and psychometric cut off points used to apply Petersen's MCI criteria.

1.2 Characterisation and differentiation of early AD and aMCI from normal ageing and depression using neuropsychological tasks

1.2.1 Cognitive Screening

Clinician surveys indicate that the Mini Mental State Examination (Folstein et al. 1975) is the most commonly used cognitive screening instrument in clinical practice (Shulman et al. 2006). Despite this, its ability to differentiate MCI sufferers from healthy and depressed elderly controls is not well established. Significant differences in the MMSE scores of these patient groups have been reported inconsistently and where they do exist, would appear to have little clinical meaning, ranging in magnitude from a minimum of less than one scale point (Ravaglia et al. 2005) to a maximum of just under 2 points (Slavin et al. 2007). As a majority of MCI patients score above the commonly used MMSE cut-offs 24/30 and 26/30, there is a considerable overlap in the scores of patients with MCI and age matched healthy controls. Much higher cut off scores, that fall within what is typically viewed as a normal range i.e. 28/30, are required to achieve an adequate level of sensitivity to combined groups of highly educated dementia and MCI sufferers (O'Bryant et al. 2008). As a result, the sensitivity of the MMSE to MCI is low with few exceptions (Callahan et al. 2002), ranging between 1% (Sager et al. 2006) and 49% (Ravaglia et al. 2005).

Longitudinal studies examining the prognostic value of MMSE scores in MCI have either found no difference in the baseline MMSE scores of well individuals who later develop MCI

(Fox et al. 1998;Meyer et al. 2002;Tang-Wai et al. 2003) and MCI patients who develop AD (Meyer et al. 2002), or they have reported statistically significant differences between non-converters and converters to dementia that are of an insufficient magnitude to be of clinical utility at an individual patient level (Aharonson et al. 2007;Amieva et al. 2005;Marcos et al. 2006).

The Addenbrookes Cognitive Examination (ACE; (Mathuranath et al. 2000) is a more extensive cognitive screening instrument, incorporating all of the items from the MMSE. This screening measure was more recently developed with the remit of improving early detection and differential diagnosis of dementia. Each of the primary domains of cognitive functioning (Lezak 2004) are sampled, and contribute to an overall score out of 100.

The ACE has become an increasingly popular cognitive assessment tool in both clinical (Alladi et al. 2006;Bak et al. 2005;Dudas et al. 2005a;Galton et al. 2005;Larner 2005;Larner 2006) and research practice (Clague et al. 2005;Dudas et al. 2005b;Estevez-Gonzalez et al. 2004;Thompson et al. 2002) within the UK. It appears to be sensitive to a relatively broad range of dementia presentations (Bak et al. 2005;Mathuranath et al. 2000) as well as to MCI (Alladi et al. 2006;Dudas et al. 2005a;Lonie et al. 2008). In an initial validation study comprising 115 patients with dementia and 124 age and education matched normal controls, a cut off score of 83/100 on the ACE showed higher sensitivity, specificity and positive predictive power for dementia than the MMSE alone (at cut offs of 27 and 24/30). Other studies have documented the test's specificity against major depression (Dudas et al. 2005a) and predictive validity in questionable dementia sufferers (Galton et al. 2005).

Sensitivity of the ACE to MCI and its specificity against affective disorders has not been replicated outside the author's (Mathuranath et al. 2000) research group. Furthermore, the prognostic power of the ACE in isolation, as it relates to MCI (Petersen et al. 1999) remains to be determined. It is conceivable that scores below the suggested dementia cut off points of 83 or 88/100 might provide a good indication of the likelihood of progression to dementia among MCI sufferers. If so, this would equip the clinician with a relatively quick and easy means of determining the likely prognosis for individual members of this patient group.

On the basis of the above, it was hypothesized that 1) in comparison with the MMSE, the more extensive ACE would better differentiate between healthy elderly controls and patients with MCI. It was also anticipated that 2) the baseline ACE score of the MCI participants would be a significant predictor of longer-term outcome, with lower baseline scores posing an elevated risk of dementia.

1.2.2 Episodic Memory

Since criteria for MCI were initially proposed (Petersen et al. 1999), a number of studies have sought to characterise the more detailed neuropsychology of this patient group. By definition, assessment of episodic memory functioning has routinely formed part of these studies, and the largest effect sizes, (where reported), are typically found in association with episodic memory measures (Alladi et al. 2006;Loewenstein et al. 2007b;Lonie et al. 2008). MCI research criteria proposed by Grundman et al (Grundman et al. 2004), specifying the use of a 1.5sd or more below age norms cut off point on the now outdated Logical Memory subtest from the WMS-R (Wechsler 1987) to establish objective evidence of episodic memory impairment, have not been universally adopted. Instead, a wide range of paradigms are being used to assess recent memory functioning in MCI.

Episodic memory paradigms can be categorised on a number of levels including (a) the nature of the material that is to be remembered i.e. verbal, visual or spatial (b) the extent to which cues are provided at recall i.e. free recall (without any cueing) or recognition (c) the number of learning trials provided (d) the length of time between encoding and recall i.e. immediate vs. delayed recall and (e) the nature of the memory to be formed i.e. an association between two or more items or a list of unrelated items. The paradigm that is employed, together with the specific measure that is chosen to assess episodic memory will determine (1) which component(s) of memory processing are evaluated (i.e. encoding, storage or retrieval) and (2) the likelihood of identifying an existing memory deficit (i.e. the test's sensitivity).

It remains unclear which type of memory paradigm and which specific neuropsychological memory measures are best suited to assessing the episodic memory failure in MCI and pre-clinical dementia, and differentiating these patient

groups from the healthy and depressed elderly (Lowndes and Savage 2007). Whilst a large number of studies have demonstrated impairments of delayed free recall in both MCI and AD (Archer et al. 2006;Arnaiz et al. 2004;Belleville et al. 2007;Bennett et al. 2006;Clague et al. 2005;Hudon et al. 2006;Kalbe et al. 2005;Nordahl et al. 2005;Perrotin et al. 2007;Riberio et al. 2006;Rose et al. 2006;Schrijnemaekers et al. 2006) deficits on such measures are not necessarily specific to MCI or pre-clinical AD and are also seen in association with depression and anxiety (Fossati et al. 2004;Swainson et al. 2001) and other non-AD forms of dementia (Salmon and Bondi 2009). As a consequence, Dierckx et al (Dierckx et al. 2007;Dierckx and Engelborghs 2007) have proposed the use of delayed cued recall or recognition measures, assuming that reduced reliance on executive aspects of memory performance results in greater specificity for MCI as it represents pre-clinical AD. In the latter study, a cued recall screening test gave 83% sensitivity and 85% specificity in discriminating individuals with AD from individuals with depression, however whilst the specificity held (85%), the test's sensitivity was seen to decrease (58%) for the comparable MCI vs. depression discrimination.

Relatively few cross-sectional studies have examined the specificity of poor episodic memory performance to MCI, and more importantly, pre-clinical AD. Several other studies that have examined this issue report the ability of cued recall measures to correctly classify subjects in accordance with their diagnostic group i.e. depression, mild cognitive impairment, normal control or early AD (Dierckx and Engelborghs 2007;Ritter et al. 2006;Swainson et al. 2001). In the latter studies, MCI patients not only exhibited deficits on the task as a whole, but were also readily divisible in terms of those whose performance mirrored the AD as opposed to the combined depression & healthy control groups. These findings suggest that cued recall measures may be a particularly useful means of delineating episodic memory impairment in MCI, as failure on such measures is likely to represent an underlying pre-clinical AD process as opposed to depression or other non-progressive psychiatric conditions.

Whilst several authors have reported impaired cued recall performance in MCI (Barbeau et al. 2004;De Jager et al. 2003;Dierckx and Engelborghs 2007;Greenaway et al. 2006;Ritter et al. 2006), others have not (Arnaiz et al. 2000;Crowell et al. 2002;Dudas et al. 2005b;Godbolt et al. 2005;Hudon et al. 2006;Westerberg et al. 2006). Arnaiz et al (Arnaiz et al. 2000) failed to find any significant difference in the performance of MCI patients and healthy age and

education matched controls on a measure of cued story recall, although deficits were apparent on the free recall components of this task. Similarly, Crowell et al (Crowell et al. 2002) found that one quarter of their MCI sample performed at ceiling levels on the recognition trial of the cognitive component of the Consortium to Establish a Registry for Alzheimer's Disease (CERAD). In another study, gist recognition memory, as defined by the number of semantically related false positive errors, was comparable among groups of MCI sufferers and age/education matched controls (Hudon et al. 2006). The dissociation between intact performance on forced choice recognition paradigms and impaired performance on yes/no recognition paradigms in MCI observed by Westerberg and colleagues (Westerberg et al. 2006) suggests that early neuropathological changes in the hippocampal and entorhinal cortex may give rise to recollection but not familiarity failure. Such findings raise uncertainty as to whether recognition paradigms are adequately sensitive to detect episodic memory failure in MCI and pre-clinical AD.

On the basis of previous findings of intact delayed recognition performance among aMCI sufferers (Arnaiz et al. 2000;Crowell et al. 2002;Dierckx and Engelborghs 2007;Dudas et al. 2005b;Godbolt et al. 2005;Hudon et al. 2006;Westerberg et al. 2006) together with the comparatively high general levels of cognitive functioning characterising our early AD and MCI participant groups, it was hypothesized that cued recall measures of episodic memory functioning would lack sensitivity to episodic memory failure in MCI and early AD whilst retaining adequate specificity (against depression and normal ageing).

Much of the selection of episodic memory measures in the assessment of MCI has been based on 1) availability of testing materials, 2) familiarity of the administrator with tests, and 3) ease of administration, rather than knowledge of the earliest sites of neuropathological change in pre-clinical AD and of which aspects of episodic memory function are likely to show compromise as a result (Lowndes and Savage 2007). The hippocampal and wider medial temporal lobe structures are known to form a vital component of the neuroanatomical circuitry that underlies the acquisition and consolidation of new information (Squire 1992). Consequently, the primary impairment in typical early AD is a failure of the encoding/consolidation process, as a result of medial temporal lobe (MTL) pathology.

The most widely accepted theory of MTL functioning is that it receives afferent information and binds it together to encode the complex relational structure of personal experiences (Cohen et al. 1999;Wallenstein et al. 1998). This includes the formation of conjunctions between spatial, temporal, perceptual, semantic, and affective information (Cohen and Eichenbaum 1993). Episodic memory measures such as the Paired Associate Learning Subtest from the Cambridge Automated Neuropsychological Touch Screen Assessment Battery (CANTAB (Robbins et al. 1994)) that require pairing of information across different modalities, have consequently been identified as important paradigms in the differential diagnosis of MCI as it represents pre-clinical AD.

The Paired Associate Learning (PAL) subtest from the CANTAB is known to be sensitive to episodic memory failure in early and pre-clinical Alzheimer's disease (Ahmed et al. 2008b;Fowler et al. 1995;Lee et al. 2003;O'Connell et al. 2004;Swainson et al. 2001) and objective memory complaints (Cargin et al. 2006;DeJager et al. 2002). Several studies have also reported that the PAL shows good specificity to early AD against depression (Swainson et al. 2001) and non-AD forms of dementia (Lee et al. 2003).

Three studies have reported on the predictive validity of the PAL subtest from the CANTAB Battery (Ahmed et al. 2008b;Blackwell et al. 2004;Fowler et al. 2002). Fowler et al (Fowler et al. 1997;O'Connell et al. 2004) observed a significant deterioration in PAL scores across an initial 6 month period in a group of 9/21 questionable dementia sufferers, all of whom went on to develop AD over the course of the proceeding 2 years. The number of errors made at the 6-pattern stage of the PAL subtest from the CANTAB was reportedly the most predictive (of a group of 7 episodic memory measures) in determining diagnostic outcome at 32 months in a further study (Blackwell et al. 2004). In a more recent study, Ahmed et al (Ahmed et al. 2008b) found that aMCI sufferers made a significantly greater number of errors than age/education matched controls at the 6 box stage of PAL at baseline. However, no differences were noted in the error rates of those MCI patients who developed dementia across the following 12 months and those who did not. Applying a cut off score of 14 errors at the 6 box level of PAL gave perfect sensitivity and negative predictive values at the expense of considerably lower (i.e. 55%) levels of specificity and positive predictive values (PPV). By combining the PAL error score with the Total ACE score, perfect sensitivity to pre-clinical AD was retained in the context of much improved levels of specificity (i.e. 82%; PPV=78%). The latter study has several limitations, including the very small number of

aMCI patients (n=18) followed up, and the short (i.e. 12 month) duration of follow-up, and would therefore benefit from replication with larger aMCI sample sizes and longer follow-up periods.

Whilst the above findings offer some preliminary hope of reaching a means of identifying AD in a pre-clinical phase, only one longitudinal study using the PAL subtest from the CANTAB battery, wherein MCI has been defined according to current criteria (Petersen et al. 1999), has been conducted (Ahmed et al. 2008b). As noted above, this study comprised a very small number of aMCI patients who were followed up for a relatively short time period (i.e. 12 months). Furthermore, the specificity of PAL against depression (Swainson et al. 2001) has not been replicated. Lastly, the limited follow-up periods of between 1-3 years have not allowed for determination of prognosis across longer time periods, which is a necessity in view of our knowledge of the relatively long i.e. up to 9 year prodromal AD phase (Amieva et al. 2005;Fox et al. 1998;Hodges et al. 2006).

For the purposes of the current study, it was predicted that 1) the PAL subtest of the CANTAB Battery would better discriminate between normal healthy elderly controls, elderly persons with depressive symptoms and AD and MCI than the more traditional paper and pencil tests of episodic memory function. It was further hypothesized that 2) the number of errors made at the 6 box level of the PAL at baseline would be a significant predictor of progression to dementia in aMCI.

Memory impairment on a background of preserved other areas of cognitive functioning forms the core for Petersen's original MCI criteria. Despite this, MCI subjects with isolated episodic memory deficits appear to comprise a relatively small proportion of the wider MCI group (Alladi et al. 2006;Loewenstein et al. 2006a;Lonie et al. 2008;Nordlund et al. 2005). Studies employing comprehensive neuropsychological batteries have consistently revealed evidence of additional impairments, in the cognitive domains of semantic memory and executive functioning, and less consistently those of psychomotor processing speed and visuospatial function (Alladi et al. 2006;Economou et al. 2007;Gualtieri and Johnson 2005;Loewenstein et al. 2006a;Loewenstein et al. 2007b;Lonie et al. 2008;Nordlund et al. 2005;Riberio et al. 2006).

Where a range of measures from each of the latter domains are employed, the presence or absence of a cognitive deficit, relative to the performance of age matched controls, is seen to vary in accordance with the specific measure that is used (Duong et al. 2006) and the general level of functioning of the MCI group. Despite this, relatively few studies have sought to compare the sensitivities of non episodic memory measures in MCI (for exceptions see (Adlam et al. 2006;Ahmed et al. 2008a;Belleville et al. 2007;Joubert et al. 2008;Murphy et al. 2006;Stockholm et al. 2006;Vogel et al. 2005), with even fewer simultaneously examining the specificity of non-memory measures to MCI. Such information is important in determining the optimal means of defining MCI criteria as it represents preclinical dementia.

1.2.3 Semantic Memory

1.2.3.1 Object knowledge

Semantic memory refers to one's store of knowledge about worldly objects, their properties, functions and associations. Semantic memory can be differentiated from episodic memory in that semantic memories are retained independently of the context in which they were learnt. Observations of impaired performance on measures of semantic memory function in both MCI (Nutter-Upham et al. 2008) and preclinical AD (Auriacombe et al. 2006) sufferers, are consistent with our knowledge of the early spread of Alzheimer pathology from the medial aspects of the temporal lobe laterally into the temporal neocortex (Braak and Braak 1991).

A variety of paradigms have been adopted to assess semantic memory function. Among these, measures of verbal fluency and confrontation naming are quick and easy to administer and among the most widely used measures of semantic memory function in the assessment of MCI. Both early AD and MCI are associated with impaired performance on measures of confrontation naming ability (Adlam et al. 2006;Ahmed et al. 2008a;Duong et al. 2006;Joubert et al. 2008) and verbal, in particular, category fluency (Henry et al. 2004;Murphy et al. 2006).

The Graded Naming (GNT; (McKenna and Warrington 1980)) and Boston Naming Tests (BNT; (Kaplan et al. 1983)) are among the most popular standardised measures of confrontation naming ability. To date, selection of confrontation naming measures in the

assessment of MCI has been governed by the availability of local normative data, with UK based studies having typically employed the GNT (Alladi et al. 2006;Blackwell et al. 2004;Swainson et al. 2001), where American based research has favoured the BNT (Loewenstein et al. 2006a;Petersen et al. 1999). Both measures were initially validated using aphasic patient groups without mention of inclusion of elderly patients for whom aphasia has arisen within the context of a dementia. Indeed, the validation sample for the 60-item version of the BNT did not include subjects over the age of 59 years. Both measures have subsequently been shown to discriminate between healthy elderly and mildly impaired AD patients (Chosak 2000;Thompson et al. 2002) and there is also evidence to suggest that even moderate levels of depression have little effect on the BNT performance of elderly subjects (Spreen and Strauss 1998).

Several recent studies have reported significantly poorer performances of MCI patients on the GNT relative to healthy age matched controls (Ahmed et al. 2008a;Alladi et al. 2006;Dudas et al. 2005b). However, earlier findings are somewhat incongruous with these results. Godbolt et al., (Godbolt et al. 2004), failed to find any impairment in performance on the GNT in a group of 19 early onset autosomal dominant familial AD sufferers in their pre-symptomatic and symptomatic (pre-diagnostic) phases. Similarly, Perry & Watson (Perry et al. 2000) failed to find any significant difference in the GNT performances of patients with minimal stage AD and age matched controls. Furthermore, Swainson et al (Swainson et al. 2001) reported that performance on the GNT failed to differentiate a group of Questionable Dementia sufferers (QD) from a combined group of depressive and healthy age matched subjects.

No study to date has reported on the prognostic utility of poor performance on this task in MCI sufferers who fulfil Petersen's MCI criteria, although De Jager et al, (De Jager et al. 2005;De Jager and Budge 2005) found that impaired GNT performance at baseline was indicative of a persisting MCI diagnosis across a 4 year period. Findings regarding the prognostic utility of baseline GNT scores in QD sufferers are mixed. In one study, baseline GNT score was found to be a significant predictor of future conversion to dementia (Blackwell et al. 2004), whereas Thompson et al (Thompson et al. 2002) observed unimpaired baseline GNT performance in a relatively small number (i.e. 6/7) of QD sufferers who went on to receive a diagnosis of AD within a two year period. Finally, a 6 year longitudinal follow-up of patients at risk of developing early onset autosomal dominant

familial AD also failed to find deficits in the GNT performances of future converters either on initial / baseline testing or at the point at which a clinical diagnosis of AD is reached (Fox et al. 1998).

Similarly Albert et al (Albert et al. 2001) found no significant difference in the baseline BNT performance of control subjects and MCI patients who went on to develop AD. Studies comparing BNT performance across groups of MCI or QD patients and healthy age matched controls have also given mixed results, with some reporting significant differences in favour of controls (Loewenstein et al. 2007b;Petersen et al. 1999) and others reporting no group differences (Balthazar et al. 2007;Greenaway et al. 2006).

Despite mounting evidence highlighting the importance of assessing semantic memory function in MCI and possible early AD cases (Hodges and Patterson 1995;Lam et al. 2006;Mickes et al. 2007), no direct comparison of the relative sensitivities of these two confrontation naming measures to MCI, preclinical and early AD has been undertaken. Furthermore, their specificity to MCI as distinct from the cognitive effects of depression, and the prognostic validity of the GNT in an MCI sample defined in accordance with Petersen's criteria has not been studied.

Within the present series of studies we therefore sought to 1) compare the relative sensitivities of the BNT and GNT to aMCI and early AD as well as 2) their specificities against the normal aging process and depressive symptoms and to 3) determine the prognostic significance of these measures in relation to MCI as defined by Petersen. In view of the absence of any preceeding literature in the first case and the contradictory nature of existing prognostic findings relating to QD and preclinical AD, no directional hypotheses were made.

1.2.3.2 Person Specific Knowledge

More recent research has examined the performance of early AD and MCI's patient's on measures of person specific semantic knowledge, as assessed by famous face naming tasks. There is accumulating evidence to suggest that measures of face naming ability may be more sensitive to early semantic memory compromise in AD and MCI than are object naming

measures (Ahmed et al. 2008a;Clague et al. 2005;Dudas et al. 2005b;Estevez-Gonzalez et al. 2004;Thompson et al. 2002;Vogel et al. 2005). Furthermore, preliminary findings would suggest that performance on measures of face naming ability may be of value in distinguishing between MCI subjects who are likely to progress to dementia over the course of the following 1-2 year period and those who do not. Thompson et al (Thompson et al. 2002) reported that of 7 from 28 Questionable Dementia of the Alzheimer Type (QDAT) patients who progressed to dementia over a 1-2 year time interval, 6 of these exhibited impaired face naming ability on initial testing. This compared with intact face naming performance for 17 of the 21 QDAT who did not progress to dementia.

This pattern of findings (i.e. face < object naming) has recently been replicated using famous buildings (Ahmed et al. 2008a) and public event (Joubert et al. 2008) recall tasks implying that representation or access to knowledge of unique semantic exemplars (as in the case of famous faces, buildings and public events) may be more vulnerable to early Alzheimer pathology than representation or access to knowledge of objects, where multiple representations of a single item exist within the lexicon.

Findings of superior sensitivity of face over object naming tasks to early semantic memory deficits in AD and MCI have been replicated on one occasion only outside the Graded Faces Naming Test (Hodges et al. 1993) developer's research group. There are no reports pertaining to the specificity of poor performance on tests of face naming ability against depression or other psychiatric or neurological conditions that frequently co-exist in older adult populations with memory complaints. Of the two studies that have reported on the prognostic utility of face naming measures in MCI, follow-up periods for each have been short (i.e. 1-2 years). Furthermore, different famous face naming tasks were employed in each of these studies. In both studies, the research group comprised the author of the face naming test that was employed. The findings therefore require replication and extension to determine specificity against depression and long-term (i.e. > 2 years) prognostic utility.

For the purposes of the present study, it was hypothesized that 1) the GFT, as a person specific naming task, would be more sensitive than both the BNT and GNT to semantic memory failure in MCI and early AD and that 2) baseline performance on this measure would contribute significantly to the prediction of conversion from aMCI to dementia.

1.2.4 Verbal Fluency

Verbal fluency tasks are an equally popular means of assessing semantic memory function in early AD and MCI patients. As they are brief to administer and easy to score, fluency measures frequently form part of bedside cognitive screening instruments and evaluations. A recent meta-analysis demonstrated that AD patients are significantly more impaired on measures of semantic than lexical fluency (Henry et al. 2004). This pattern of impairment of verbal fluency measures is qualitatively distinct from the usual finding of superior semantic fluency in healthy controls (Spreen and Strauss 1998). As the category task is thought to rely more heavily on access to representations of semantic concepts than the letter task, the pattern of findings in AD is presumed to reflect degradation in the structure, content or activation of the semantic memory system (Auriacombe et al. 2006;Jefferies and Lambon Ralph 2006;Jones et al. 2006).

Patients in the pre-clinical stages of AD exhibit a semantic fluency deficit, at a time when lexical fluency performance remains intact (Auriacombe et al. 2006;Beatty et al. 2002;Swainson et al. 2001). Similarly, patients who fulfil criteria for Amnestic Mild Cognitive Impairment (aMCI; (Grundman et al. 2004;Petersen et al. 1999)) generate fewer words from a specified category than do age matched controls. In contrast, they perform at normal levels on lexical fluency tasks (Alladi et al. 2006;Dudas et al. 2005b;Lonie et al. 2009a;Murphy et al. 2006).

A pattern of worse semantic than lexical fluency has also been reported in patients with depression (Christensen et al. 1997;Zakzanis et al. 1998), although a more recent review suggests equal impairment of performance across the two fluency tasks, thought to reflect a generalised reduction in processing speed (Henry and Crawford 2005).

If semantic and lexical fluency discrepancy scores are abnormal in some patients with aMCI, their magnitude and direction may prove helpful in diagnosis and/or prognosis. As an individually calibrated marker of performance, the direction of the discrepancy would have the advantage of being free from the need for age, gender, education or IQ-dependent cut off values, which require a sizable normative comparison group.

Furthermore, if depressive symptoms were associated with equivalent reductions in lexical and semantic task performance (as the processing speed account would predict), then fluency discrepancy scores might also be of value in distinguishing between depressive and early Alzheimer related cognitive impairment.

One study has reported on the use of semantic-lexical fluency discrepancy scores in detecting MCI. Murphy et al (Murphy et al. 2006) demonstrated a progressive advantage (controls > aMCI > AD) in semantic, relative to phonemic fluency, with difference scores between tasks distinguishing each group from the others with medium to large effect sizes ranging from 0.49 to 1.07. However, this study lacked the important inclusion of a depressive control group matched for general level of cognitive functioning and support needs. The differential diagnosis of depression and early stage dementia in the elderly is notoriously difficult (Lezak 2004), and it is for this reason important to establish the extent to which poor performance on a cognitive measure is attributable to an underlying neurodegenerative disease process as opposed to an affective disorder. As such, it remains uncertain to what degree the pattern of lexical>semantic fluency performance is specific to aMCI and AD and whether a similar pattern of findings might generalise to the use of alternative letters, (in particular those that already form part of widely used dementia screening instruments such as the ACE). The prognostic significance of this pattern of fluency performance has not been studied.

It was hypothesized that fluency discrepancy scores would be a sensitive and specific means of differentiating MCI and early AD patients from healthy elderly controls and elderly controls with depressive symptoms.

1.2.5 Executive Function

Executive function is an umbrella term describing a number of different cognitive processes (i.e. abstract reasoning, divided attention, inhibition, mental flexibility, self-monitoring, planning, problem solving, working memory) that serve a 'global processing control function' (Attix and Welsh-Bohmer 2006) facilitating pre-meditated, purposeful and goal directed behaviour. Executive functions are thought to be mediated in part by structures

within the frontal brain cortex and damage to this region has been associated with impairment in each of the processes outlined above (Lezak 2004).

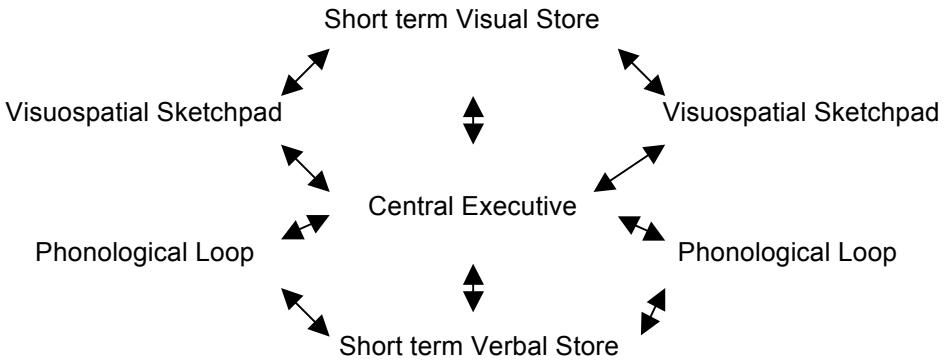
A number of studies have demonstrated the presence of deficits of an executive nature in preclinical AD (Backman et al. 2005) and early AD (Egerhazi et al. 2007;Perry et al. 2000;Stockholm et al. 2006). MCI patients have been shown to perform significantly less well than their age peers on a range of standardised executive measures including the Wisconsin Card Sorting Test (WCST; (Clague et al. 2005), Similarities subtest from the WAIS-III (Loewenstein et al. 2007b), The Spatial working memory subtest from the CANTAB battery (Collie et al. 2002;Egerhazi et al. 2007), digit span subtest from the WAIS-III; forwards (Dudas et al. 2005b;Lu et al. 2005) and backwards (Crowell et al. 2002;Lu et al. 2005), spatial span subtest from the WAIS-III; forwards and backwards (Levinoff et al. 2006) and the Brown-Peterson procedure (Belleville et al. 2007). By contrast, performance on a range of measures assessing aspects of planning, estimation and problem solving ability has failed to discriminate between MCI sufferers and age matched controls (Collie et al. 2002;Egerhazi et al. 2007;Levinoff et al. 2006;Nordlund et al. 2007;Swainson et al. 2001).

As is true for other domains of cognition, there is some preliminary evidence to suggest that the ability to detect early compromise of an executive nature in MCI and preclinical AD varies in accordance with the aspect of executive functioning that is assessed, and the specific measure that is employed. Belleville et al (Belleville et al. 2007) demonstrated that divided attention was impaired in MCI, whereas this group performed at similar levels to age matched controls on measures tapping other aspects of attention control (i.e. mental manipulation and inhibition). As the Petersen-Brown task that was employed to assess divided attention also required patients to retain three consonants across a 30 second delay period, and MCI patients were found to be impaired on this task without a divided attention component, it was not clear whether their reduced performance was reflective of defective divided attention or rather short term memory. Similarly, an impairment of response inhibition in MCI patients documented using an experimental paradigm (The flanker task; (Wylie et al. 2007) has not been observed on several standardized clinical measures of response inhibition such as the Stroop; (Belleville et al. 2007;Hudon et al. 2006;Kramer et al. 2006;Nordlund et al. 2007) and the Hayling; (Belleville et al. 2007).

The most consistent deficits of an executive nature in MCI have been reported in association with performance on Part B of the Trail Making Test (TMT Part B; (Reitan 1985) (Alladi et al. 2006;Archer et al. 2006;Arnaiz and Almkvist 2003;Crowell et al. 2002;Loewenstein et al. 2007b). These findings imply compromise to complex attentional functions such as dual processing, working memory and the shifting and division of attention in the pre-clinical phase of AD. Deficits of this nature are a well established feature of early AD (Della Sala and Logie 2001;Parasuraman and Haxby 1993) and at least one study suggests that poor performance on Part B of the TMT is a significant predictor of time to progression from MCI to a clinical diagnosis of AD (Blacker et al. 2007). In a recent meta-analysis however, Herrmann et al (Herrmann et al. 2007) concluded that executive deficits, as assessed by TMTB among other measures, are also common among elderly depressed patients, particularly those suffering with a late onset depression. As such, it would seem important to establish the specificity of poor performance on part B of the TMT to MCI prior to determining the predictive capabilities of this measure.

The dual task measure similarly assesses the ability to co-ordinate the simultaneous performance of multiple tasks and is purported to be a sensitive marker of executive failure in early AD (Baddeley et al. 1986;Collette et al. 1999;Della Sala et al. 1995;Greene et al. 1995). The measure was developed by Baddeley and colleagues in order to evaluate the central executive component of working memory (Baddeley and Hitch 1974). Figure 1.4 depicts the model of working memory. Within this model, the role of the central executive is to coordinate the simultaneous operation of the other components. In doing so, the central executive facilitates temporary retention of new information in a short term store and the maintenance of such information via mental rehearsal so that mental manipulation can be carried out in an online manner. By comparing performance in a dual task situation to performance on the same tasks performed alone, it is suggested that any drop in performance can be attributed to a failure of the central executive.

Figure 1.4 Baddeley and Hitch (1974) Model of Working Memory



Dual task deficits have been reported in association with AD (Baddeley et al. 2001;Logie et al. 2004) although not consistently (Greene et al. 1995;Perry et al. 2000), and in elderly patient's without a dementia diagnosis with Dementia Rating Scale Scores falling below 123/144 (where 123 is the suggested cut off point for dementia) (Holtzer et al. 2004). As performance on the dual task can be seen to vary in a qualitative as well as quantitative sense (i.e. no decrement vs. decrement), it is potentially well suited for use as a tool in early differential diagnosis of AD as has been described as a 'viable and fruitful approach to the development of clinical tools' (Della Sala and Logie 2001). Furthermore, dual-task performance is reportedly insensitive to the normal ageing process, in so far as the individual components of the overall task rely on different modalities (i.e. visual vs. auditory) and that these single components are adjusted so that difficulty levels are equal across comparison groups (Baddeley et al. 1986;Della Sala and Logie 2001).

Studies that have examined the performance of depressed patients on dual task measures, without adjusting the levels of performance on the component tasks performed singly to those of controls, show mixed results, some reporting a dual task decrement in association with depression (Arnett et al. 1999;Lemelin and Baruch 1998) and others reporting the opposite pattern of findings i.e. a greater dual task decrement in controls than patients with depression (Thomas et al. 1999). A more recent study, in which levels of performance on the component parts of the dual task were matched across groups, found a significantly greater dual performance deficit in depressed patients that was seen to persist with remission (Nebes et al. 2001).

No study to date has reported on the dual task performance of patient's who fulfil Petersen's criteria for MCI, or examined the longitudinal course of dual task performance in this patient group. As noted above, there are mixed findings relating to the sensitivity of the Dual Task to executive compromise in early AD and depression. The measure's potential utility in the early identification of pre-clinical AD therefore remains uncertain.

It was hypothesized that measures of divided attention and dual task performance but not lexical fluency would be sensitive to early executive failure in MCI and AD

1.2.6 Visuospatial Function and Processing Speed

Alongside attention, episodic, semantic, working memory and executive functioning, visuo-construction and processing speed comprise the remaining two sub-domains of cognitive processing as established by neuropsychological profiles in different forms of dementia (Lezak 2004). Deficits in the latter two domains have also been reported in association with amnestic MCI, albeit, with considerably less frequency than other cognitive domains. Such observations would fit well with what is known of the temporal sequence of neuropathological change in AD, where neurofibrillary tangles are seen to accumulate within the medial and lateral portions of the temporal lobe prior to spreading to posterior association cortices (Braak and Braak 1991).

In the assessment of MCI, visuo-spatial and constructional abilities have been most frequently assessed with copied drawing tasks, such as the Rey Complex Figure Test (Rey 1941). Several authors have reported significantly lower mean performances of MCI patients relative to healthy elderly controls on the copy component of the RCFT (Alladi et al. 2006;Archer et al. 2006). There are a number of other formal neuropsychological measures comprising one or more subtests that have been designed to assess visuo-spatial, visuo-constructional and visuo-perceptual abilities. A majority of studies employing measures of visuospatial/constructional and perceptual ability have failed to reveal significant differences in the performance of MCI patients and age matched controls on the RCFT-copy (Adlam et al. 2006;Dudas et al. 2005b;Kalbe et al. 2005;Lambon et al. 2003;Nordlund et al. 2007) the Block Design subtest from the Wechsler Adult Intelligence Scale – III (WAIS-III;(Wechsler 1997) (Bennet 2006; Greenaway 2006; Levinoff 2006; Nordlund 2007), the Visual Object and Space Perception Battery (VOSP;(Warrington and James 1991) (Adlam et al. 2006;Archer et al. 2006;Clague et al. 2005;Dudas et al. 2005b;Nordlund et al. 2007) and the constructional praxis component of the Consortium to Establish a Registry for Alzheimer's Disease (CERAD; (Morris et al. 1989) (Bennett et al. 2006;Crowell et al. 2002).

Such observations are in keeping with the findings of a recent meta-analytic review of cognitive impairment in pre-clinical AD (Backman et al. 2005), where smaller differences in the performances of pre-clinical AD and healthy elderly on measures of visuo-spatial ability were observed relative to tests of episodic memory, global cognitive function and processing

speed. There is relatively little information addressing the prognostic validity of impaired visuo-spatial and constructional ability in MCI. Some authors report baseline visuo-perceptual (Marcos et al. 2006) visuo-constructional (Artero et al. 2003) performances as significant predictors of future conversion to AD and others have found no association between visuo-spatial performance and risk of AD (Aggarwal et al. 2005;Albert et al. 2007;Albert et al. 2001). The heterogeneous nature of AD presentations is well established (Caine and Hodges 2001), and it is possible that the assessment of visuo-spatial function may be of greater importance in younger MCI patients given the evidence to suggest that atypical (i.e. visual) presentations of AD are more common among early onset sufferers (Black 1996).

Whilst early AD, pre-clinical AD, and aMCI clinical presentations are not typically associated with the reductions in psychomotor processing speed that accompany the subcortical dementias, recent meta-analytic findings indicate that pre-clinical AD patients (Backman et al. 2005) process information at a slower speed than their age matched contemporaries. Neuropsychological assessment of processing speed most frequently comprises a timed task requiring both motor and cognitive input (i.e. a visuomotor processing speed task) examples of which include, the Digit Symbol subtest from the WAIS-III (Wechsler 1997), Part A of the Trail Making Test (TMTA; (Reitan 1985)) and the Symbol Digit Modality Test (SDMT; (Smith 1982). Of these, TMTA has been the most popular choice in assessing psychomotor speed in MCI. A majority of studies have failed to find a difference in the time taken by MCI patients and healthy elderly controls to complete this task (Archer 2006; Arnaiz 2000; Belleville 2007; Bennett 2006; Crowell 2002; Greenaway 2006; Kalbe 2005; Westerberg 2006) but see (Alladi et al. 2006;Arnaiz et al. 2000), suggesting that visuo-motor processing speed (when assessed in this manner among smaller MCI patient groups) may lack sensitivity to MCI. It was therefore hypothesized that the time taken to complete TMTA would not differ between MCI and healthy elderly control participant groups

1.3 Predicting conversion from aMCI to dementia using neuropsychological measures

Current MCI criteria do not specify the manner in which cognitive deficits should be defined. There is, as a result, scope for variability in the case definition of MCI in terms of 1)

the number of cognitive domains that are impaired 2) the number and nature of episodic memory measures that are used to detect memory impairment and 3) the severity of episodic memory impairment that is observed. Each of these factors could conceivably affect the proportion of MCI sufferers who go on to develop dementia as well as the rate at which progression to dementia occurs. A greater understanding of the influence that variability in MCI case definition has on clinical outcome would inform the use of more detailed cognitive criteria that are better able to identify a more homogenous group of pre-clinical dementia sufferers.

1.3.1 Impairments of a non-episodic memory nature as predictors of progression to dementia

Since the original MCI criteria were expanded in recognition of four subtypes (see figure 1.2 above), interest has developed in the prognostic significance of multiple vs. single domain amnestic MCI. Research has thus far focused on whether deficits in domains other than that of episodic memory function are predictive of future conversion to dementia and whether greater numbers of impaired cognitive domains at baseline are associated with an increase risk of conversion from MCI to dementia.

In relation to MCI as it is defined by Petersen (Petersen et al. 1999), poorer baseline performance on measures of processing speed (Aggarwal et al. 2005;Amieva et al. 2004;Bennett et al. 2002;Tabert et al. 2006), semantic memory function (Bennett et al. 2002;Estevez-Gonzalez et al. 2004), and less frequently visuo-perception (Marcos et al. 2006) and executive functioning (Perri et al. 2007) have been reported as significant predictors of future conversion to dementia across 2-10 year time periods. Where 'at risk' groups have been defined by some means other than Petersen's MCI criteria, semantic memory measures are the most frequently reported significant predictors of conversion to dementia (Artero et al. 2003;Galton et al. 2005;Guarch et al. 2004) behind episodic memory tasks.

By comparison, relatively few studies have addressed the influence that the number of cognitive domains impaired at baseline has on risk of conversion to dementia. In a 2 year retrospective evaluation of 48 non-demented and non-depressed elderly patients with clinical and neuropsychological evidence of memory deficits, Bozoki et al., (Bozoki et al. 2001)

found significantly higher rates of conversion to AD among the MCI patients with multi-domain amnestic (48% & 77%) as compared to single domain amnestic MCI subtype (6% & 24%) in the 2nd and 4th years of follow-up respectively. A similar pattern of findings was reported by Tabert et al., (Tabert et al. 2002) who observed a significantly greater proportion of multi-domain amnestic MCI patients (50%) convert to AD across a three year time period than patients with single domain amnestic MCI (10%). By contrast, Storandt et al., (Storandt et al. 2006) reported similar rates of progression (0.5SD/year on a psychometric composite), mean survival times to a Clinical Dementia Rating scale score of 1, and neuropathologic diagnoses of AD, among 32 individuals with single and 90 people with multiple domain amnestic MCI subtype. As such, the extent to which the co-existence of cognitive deficits outwith the domain of episodic memory places MCI sufferers at increased risk of conversion to dementia, remains uncertain.

On the basis of what is known of the topographical spread of AD pathology in the preclinical and early AD phases (Braak and Braak 1991) and the necessity for evidence of cognitive impairment in at least one non-memory domain in order to fulfil DSM-IV criteria for AD, a positive association between the number of domains of cognitive impairment, as well as the presence of semantic memory impairment in addition to that of episodic memory, and risk of conversion to dementia might be expected. It was therefore hypothesised that a greater extent of non-memory impairment in aMCI at baseline would be associated with greater risk of conversion to dementia at long term follow-up.

1.3.2 Consistency of episodic memory impairment as a predictor of progression to dementia

Several studies have documented a lack of temporal stability and poor predictive validity for conversion to AD using the original Petersen criteria (De Jager and Budge 2005;Ritchie et al. 2001), suggesting that reliance on the use of a single episodic memory measure to define objective memory deficit may give rise to unstable MCI diagnoses (Brooks et al. 2007). Reduced motivation, lapse of attention or distraction could conceivably give rise to test failure among the normal elderly, during a single evaluation. Whilst the use of combined memory test scores has been found to provide better sensitivity to MCI than single tests (De Jager and Budge 2005), uncertainty remains as to what effect consistency of impairment across multiple episodic memory tests might have on the stability of MCI classification as well as long-term diagnostic outcome.

Rotrou et al., (de Rotrou et al. 2005) followed-up a group of 130 sixty to eighty-year old healthy autonomous volunteers who undertook a comprehensive neuropsychological battery of tests (assessing aspects of language, memory and executive function) at baseline, 6 and 12 months. The authors observed a good deal of instability of MCI classification, with 48% of participants who performed more than 1.5 standard deviation below control means on one or more neuropsychological measure at baseline, no longer doing so when re-tested 12 months later. Furthermore, the MCI subjects who remained impaired at the 12 month follow-up had low scores at baseline on three tests or more, compared with one or two failed tests for those MCI participants who normalised. Twelve of thirteen participants, who failed three tests or more at baseline, remained impaired at their 12-month follow-up. The authors concluded that in normal elderly, failing three tests or more could be more predictive of pathological outcome than failing one or two tests whilst conceding that it was entirely possible that the group of unstable MCI sufferers remained at higher future risk of dementia conversion than their non-MCI counterparts. Cognitive deficits of an unstable nature have also been observed in a small group of pre-clinical AD sufferers across the early prodromal years (Hodges et al. 2006).

Insofar as one might expect the base rates of 'impaired' performance on two measures of episodic memory function among normal elderly controls to be lower than for a single measure, demonstration of impairment on multiple measures of episodic memory function in defining Petersen's MCI criteria, may allow for the maintenance of adequate levels of sensitivity to MCI whilst simultaneously reducing the risk of false positive and unstable MCI classifications.

It was therefore hypothesized that increased pervasiveness of the episodic memory impairment displayed by aMCI patients at baseline, would be associated with an elevated risk of receiving a diagnosis of dementia over time.

To summarise, performance on one or more specific neuropsychological task(s), the consistency of the episodic deficit at baseline and the extent to which impairments in other cognitive domains are present, may help to predict long-term outcome in aMCI. The relative

contribution of each of these in predicting the clinical course of aMCI remains to be determined.

2. Materials & Methods

2.1 Participants

In this thesis, 46 patients with aMCI; 20 outpatients with depressive symptoms; 24 healthy control patients and 21 patients with mild AD were studied.

2.1.1 aMCI

The 46 aMCI patients were recruited over a 4-year period (Sept 2003 – Sept 2007) from tertiary referrals to the Royal Edinburgh Hospital Neuropsychological Assessment Service and met criteria for aMCI (Petersen et al., 1999). The service operates at a tertiary referral level, with the bulk of patients referred from Lothian-wide Old Age Consultant Psychiatrists and lesser numbers from Geriatricians and Neurologists.

Thirty seven of the 46 MCI participants also underwent psychiatric evaluation prior to their clinical neuropsychology consultation. All aMCI patients in our sample had presented to their GP with memory complaints, that were (other than those for whom no family or friend was present at interview n=2) corroborated by an informant and had undergone routine medical and blood screening for dementia. Patients were referred to consultants in Old Age Psychiatry (n=71) or Medicine of Old Age (n=16) for further evaluation of their memory complaints. Each of the 87 MCI patients who were invited to participate in our study had also been referred for, and completed, a clinical neuropsychological assessment prior to study entry.

Clinical neuropsychological assessments varied in their length and content according to the patient's cognitive capabilities, willingness to participate and the nature of the referral question. Typically, assessments included a structured clinical interview followed by a measure of overall level of cognitive functioning together with one or more measures assessing episodic and semantic memory, language, visuo-spatial and executive function. There was a considerable degree of overlap in the tests employed in clinical practice and those comprising the research battery such that the baseline clinical assessment for the aMCI patient group was used as the baseline study assessment. Where necessary, outstanding tests

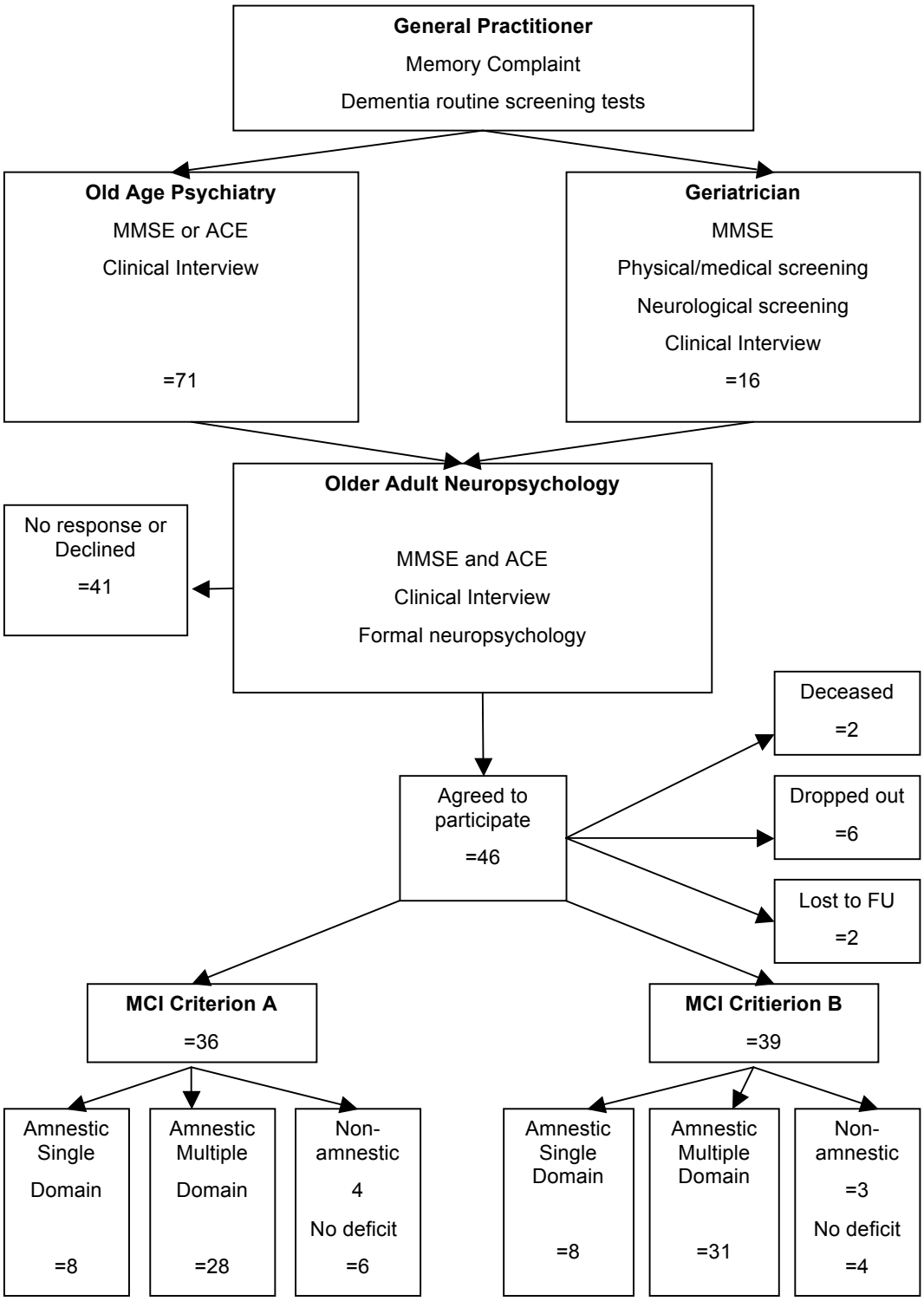
from the study battery were administered on a separate occasion, no more than 3 months post the baseline clinical assessment.

Although a cut off value of 1.5SD is commonly applied as a means of delineating cognitive impairment in MCI in the research literature, in clinical neuropsychological practice, it is desirable to demonstrate some reliability of performance across tests. In view of this, and the above average mean pre-morbid IQ estimate for the MCI group, both the severity and consistency of impaired cognitive performance at the initial clinical neuropsychological assessment were used as criterion for recruiting MCI participants. Two sets of criteria were applied in recruiting MCI participants based on their clinical neuropsychological findings. Using criterion A, cognitive impairment was defined by a performance of 1sd or more below published age means (Benedict et al. 1998;Graham 2004;Saxton et al. 2000;Spreen and Strauss 1998;Tombaugh et al. 1999;Tombaugh 2004;Warrington 1997) on two or more measures assessing a single cognitive domain (see Table 2.1 for a breakdown of measures in accordance with the cognitive domains broadly assessed). For criterion B, cognitive impairment was defined in terms of a performance 1.5sd or more below published normative age means (Benedict et al. 1998;Graham 2004;Saxton et al. 2000;Spreen and Strauss 1998;Tombaugh et al. 1999;Tombaugh 2004;Warrington 1997) on at least one test.

All MCI patients who were invited to take part in the study met criterion A (in relation to published age norms (Benedict et al. 1998;Graham 2004;Saxton et al. 2000;Spreen and Strauss 1998;Tombaugh et al. 1999;Tombaugh 2004;Warrington 1997) at their initial clinical neuropsychological consultation. A breakdown of MCI study participants in accordance with criterion A and B, on the basis of their baseline study neuropsychological performance is provided in Figure 2.1 at the end of the flowchart.

In all cases, levels of cognitive performance and / or everyday functioning, based on cognitive screening measures, clinical interview and the informed opinion of an experienced clinical neuropsychologist, were considered sufficiently preserved to prevent a diagnosis of dementia being made. Participants in whom a functional cause to cognitive complaints was suspected or established following clinical interview and or the administration of mood screening instruments, were excluded.

Figure 2.1 Recruitment flowchart and classification of MCI subtypes at baseline



Criterion A, > or = 1sd below healthy elderly age and IQ matched study control group on 2 or more cognitive measures of a single domain; Criterion B, > or = 1.5sd below healthy elderly age and IQ matched study control group on at least 1 cognitive measure of a any given domain; FU, follow-up.

Table 2.1 Numbers of persons followed up for between 1 - 5 years as a function of the MCI criterion met at baseline assessment.

Number of Follow-up assessments	Criteria A (number of participants)	Criteria B (number of participants)	'Other' (number of participants)	Total
1	2	2	0	4
2	8	9	1	18
3	13	15	3	31
4	10	10	5	25
5	3	3	1	7

Twenty-nine of 46 of the aMCI group underwent neuroimaging at some point during the course of the study. Nine subjects underwent CT scanning, 2 underwent SPECT & a further 6 underwent DTI. Four aMCI participants underwent both CT and SPECT scanning and 3 underwent CT, SPECT and DTI. Two aMCI participants underwent CT and DTI scanning with each of the remaining 3 undergoing combinations of CT, SPECT & MRI; CT, SPECT & DIT and fMRI & DTI, respectively. Amnestic MCI patients, who were exhibiting evidence of significant and predominant cerebro-vascular disease or other predominant changes of a non-atrophic nature on neuroimaging *prior* to their study participation that could account for their cognitive presentation, were excluded from the study (see Table 2.2 for details of inclusion and exclusion criteria). Seventeen aMCI participants did not undergo any form of neuroimaging making it impossible to rule out the presence of structural or non-atrophic change in these cases.

2.1.2 Early Alzheimer's Disease

Twenty one patients with a NINCDS/ADRDA (McKhann et al. 1984) diagnosis of probable AD were identified as part of a clinical audit of service referral numbers and diagnoses. All early AD patients scored 17/30 or above on the Mini Mental State Examination (MMSE; (Folstein et al. 1975) and 58/100 or above on the more comprehensive Addenbrookes Cognitive Examination (ACE; (Mathuranath et al. 2000) indicating a relatively mild disease severity. Patients with known and potentially confounding co-morbid medical, psychiatric or neurological conditions (i.e. stroke or cerebro-vascular disease, head injury, alcoholism, schizophrenia, etc.), as established by way of review of medical notes, were excluded (Table

2.2). For one of the early AD participants, MRI scanning that was carried out at a later date as part of a related MRI-DTI imaging study, confirmed the presence of a possible carotid artery occlusion. In a further case, MRI-DTI scanning conducted post recruitment phase, revealed evidence of a possible arterial venous malformation. Neither of the participants was excluded from the study, and both participants are represented in the tables, figures and analyses comprising early AD participants.

2.1.3 Healthy Elderly Controls

As healthy controls we recruited 24 spouses or carers of patients who had attended the Neuropsychological Assessment Service. Participants with known and potentially confounding co-morbid medical, psychiatric or neurological conditions (i.e. stroke or cerebrovascular disease, head injury, alcoholism, schizophrenia, etc.), as established by way of structured interview, were excluded (see Table 2.2).

2.1.4 Depressive Elderly Controls

In an attempt to ensure a control group of similar illness severity and general level of functioning to the aMCI patients, we recruited 20 outpatients with depressive symptoms from hospital outpatient clinics and day hospitals, who were receiving the same level of outpatient care as our aMCI group. All participants in this group presented with depressive symptoms, thought not to be primarily organic in nature, yet known to have effects on cognitive functioning both during illness and after recovery (Herrmann et al. 2007).

We considered matching of illness severity to be important, as in clinical practice the differentiation of severe depression and early dementia states is less problematic than separating the sequelae of the milder forms of these disorders. We included patients with a variety of disorders, as the type of depression does not appear to influence the magnitude of cognitive deficit (Christensen et al., 1997). As with the other participant groups, patients with any co-morbid medical, neurological or additional psychiatric condition with the potential to affect cognitive function were excluded by way of review of the medial notes. The mean Geriatric Depression Scale (GDS; (Yesavage et al. 1983) score for this group was 19 (SD=5.24; 95% CI= 16.55-21.45) indicating mild-moderate, clinically significant levels of depressive symptoms. Information pertaining to the date of onset of depressive illness (i.e. before or after 60 years of age/ early or late onset) and the presence or absence of deep white matter hyperintensities on neuroimaging, both of which have been found to affect the neuropsychological status of depressed elderly individuals (Steffens, 2008), was not collected. The interpretation of the neuropsychological performances of the depression

control group was limited by the absence of this information (see limitations section Chapter 7 for fuller discussion).

Table 2.2 Participant Inclusion / Exclusion Criteria

Participant Group	Inclusion Criteria	Exclusion Criteria
aMCI	Subjective complaints of memory difficulty and /or decline	Untreated/ or unsuccessfully treated depressive illness – medical notes
	Performance greater than one standard deviation below age and IQ mean on at least two measures of episodic memory function	History of Head Injury
	Functional decline very minimal or absent – structured clinical interview	History of Stroke, epilepsy or other neurological condition(s)
	Does not fulfil diagnostic clinical criteria for dementia	Medical condition(s) that might account for memory loss i.e. hypothyroidism
	>55 years of age	< 55 years of age
		Evidence of significant and predominant cerebro-vascular disease on neuroimaging
		Diagnosis of dementia
Controls	>55 years of age	<55 years of age
	Matched for age, Pre-morbid IQ, sex distribution to MCI group	Current depressive illness – self reported low mood
		Subjective cognitive complaints
		Medical or neurological condition known to be associated with cognitive impairment
Early AD	Diagnosis of AD NINCDS & DSM-IV	<55 years of age
	MMSE score 20/30 or greater	Co-morbid untreated/unsuccessfully treated depressive disorder – medical notes
	>55 years of age	
	Matched for age, Pre-morbid IQ, sex distribution to MCI and control group	Co-morbid neurological or medical condition that is known to be associated with cognitive impairment
		Significant degree of cerebrovascular change on neuroimaging
Depressive Symptoms	GDS>10	<55 years of age
	Diagnosis of Major depression	Diagnosis of dementia
	Or	Subjective cognitive complaints documented in medical notes or evident during testing
	Currently receiving hospital treatment for depressive symptoms	GDS<10
	>55 years of age	Co-morbid neurological or medical condition that is known to be associated with cognitive impairment evident in the medical notes or dudring testing
	Matched as far as possible in terms of age, pre-morbid IQ, sex distriction to aMCI, control & early AD group	
		Abnormalities on neuroimaging as evidence in the medical notes

MCI = Mild Cognitive Impairment; Pre-morbid IQ=Intelligence Quotient; AD=Alzheimer's disease; GDS= Geriatric Depression Scale; MMSE=Mini Mental State Examination.

Fourteen of twenty patients comprising the elderly depressive normative sample had a diagnosis of major depressive disorder. The remaining participants met DSM-IV criteria for Dysthymic disorder (n=4), Bipolar I Disorder (most recent episodic depressed n =1) or Anxiety Disorder (n=1). Sixteen of the 20 depressive participants were being treated with antidepressant medication at the time of their study participation. Of the remaining 4, 2 were taking anxiolytic forms of medication and two were receiving intervention of a non-pharmaceutical nature only. A summary table of diagnostic and treatment details for the depression group can be found in Table 8.1 of the Appendix.

2.2 Neuropsychological Measures and Psychometric Test Properties

All four patient groups undertook the following battery of neuropsychological tests at baseline; Addenbrookes Cognitive Examination (ACE; (Mathuranath et al. 2000)), Mini Mental State Examination (MMSE; (Folstein et al. 1975)), National Adult Reading Test (NART; (Nelson and Willison 1991)), Paired Associate Learning Test from the Cambridge Neuropsychological Test Automated Battery (CANTAB, PAL; (Morris et al. 1987)), Rey Complex Figure Test (ROCF; (Rey 1941)), Hopkins Verbal Learning Test-Revised (HVLT-R; (Brandt 1991)), Graded Naming Test (GNT; (McKenna and Warrington 1980)), Edinburgh Exemplar Naming Test (EENT; JAL)), Boston Naming Test (BNT; (Kaplan et al. 1983)), Graded Faces Test (GFT; Graham, personal communication 2004), Category fluency using the categories 'animals', 'fruits' & 'vegetables', Controlled Oral Word Association Task using the letters 'F', 'A' & 'S' (COWAT; (Spreen and Strauss 1998)), Dual-Task (Della Sala, personal communication, 2004; (Della Sala 2005)) and the Trail Making Test (TMT; (Reitan 1985)). Classification of each of these tasks in accordance with their broad associations with specific cognitive domains is provided in Table 2.3. A brief description and summary of the relevant psychometric properties for each of these measures is provided below.

Depression patients were also administered the Geriatric Depression Scale (GDS; (Yesavage et al. 1983)) as a means of ensuring they met inclusion criteria for this group. MCI patients were re-administered the neuropsychological test battery annually for the duration of their participation in the study (which ranged from a minimum of 2, to a maximum of 5 years). 16 of 24 healthy elderly control participants undertook the neuropsychological test battery on

two occasions, at varying 1 – 3.5 year intervals. At the final assessment, the primary carer (or a close friend where the latter was not available) of every MCI patient also completed a set of functional scales detailing perceived changes in their ability to carry out higher level (Farias et al. 2006) as well as more rudimentary activities of daily living and personal and self maintenance tasks (Lawton and Brody 1969). Details of the neuropsychological assessment protocol at each stage of follow-up, for all participant groups, are tabulated in Table 8.2 of the Appendix. An indication of broad groupings of neuropsychological measures in accordance with their cognitive domain coverage is provided in Table 2.3 below.

Table 2.3 Neuropsychological measures grouped in accordance with cognitive domain coverage

Cognitive Screening Measures	MMSE, ACE
Pre-morbid IQ	NART-R
Episodic Memory	PAL, RCFT, HVLT-R
Semantic Memory	GNT, EENT, BNT,GFT, Category fluency
Attention and Executive Functions	TMT A & B, COWAT, Dual Task
Functional Ability	Everyday functioning, PSMS, ADLS
Mood	GDS

2.2.1 Cognitive Screening

2.2.1.1 Mini-Mental State Examination (MMSE; (Folstein et al. 1975))

The MMSE is a brief cognitive screening measure requiring less than 10 minutes to administer. It is usually administered as a means of establishing an individual's overall level of general cognitive functioning. It is presently the most widely used cognitive screening instrument for dementia (Lezak 2004). The MMSE samples a restricted set of cognitive functions including concentration and working memory, language and praxis, orientation, memory and attention span. The measure comprises 19 questions yielding a possible total of 30 points. Cut off scores of 27/30 and 24/30 are usually applied in determining the likelihood of dementia or some other form of organic brain pathology, with variability in accordance with the age and education level of the individual patient (Lopez et al. 2005). Scores of between 21-26/30 are taken as indicative of mild cognitive impairment, 11-20/30

as indicative of moderate levels of impairment and those below 10 signify severe cognitive impairment (Folstein et al. 2001).

Scores on the MMSE are positively influenced by education and negatively by age, whereas gender appears to have little effect on MMSE performance. Pre-morbid IQ estimates have also been shown to influence MMSE score in mild and moderate AD (Starr and Lonie 2007). Some authors have reported high test-retest reliabilities of between r=0.89 and r=0.99 over 24-hour and 4 week periods for non-demented and dementia patient samples, respectively (Lezak 2004). Others have reported 95% confidence intervals for true scores as large as 10 points for an obtained score of 23/30 (Lopez et al. 2005). Despite such variability, the performance of healthy older adults is generally considered to be stable over the longer term (Lezak 2004).

2.2.1.2 Addenbrookes Cognitive Examination (ACE; (Mathuranath et al. 2000))

The ACE is a more comprehensive cognitive screening measure than the MMSE, requiring around 15-20 minutes to administer. The measure synthesises all of the existing MMSE items with components of well established neuropsychological tasks. The ACE was developed with the aim of increasing the sensitivity of existing screens to milder forms of cognitive impairment and assisting in the early differential diagnosis of the dementias. The test includes a total of 32 items covering the domains of attention/concentration, orientation, recall, memory, verbal fluency, language, and visuospatial abilities, yielding a maximum score of 100. Sub-scores may be calculated for each of the above domains. As the ACE comprises all of the MMSE items it is possible to derive a separate MMSE score simultaneously.

The ACE is sensitive to dementia using a cut off value of <88/100 (sensitivity = 93%). There is, however, no data reporting on practice effects or test re-test reliability among the healthy elderly. Total ACE scores are known to correlate very highly with MMSE scores (r=0.92(Larner 2005)). More favourable sensitivity, specificity and positive predictive values for dementia relative to non-dementia (Bier et al. 2004;Mathuranath et al. 2000) and Questionable Dementia patients who do, versus to those who don't convert (Galton et al. 2005) to dementia have been reported for the ACE as compared to the MMSE. Specificity to

dementia (spec = 96%) is reportedly similar for ACE and MMSE at cut offs of 83/100 & 27/30 respectively) but less for the ACE (42.6%) than the MMSE (53.8%) using the higher 88/100 ACE cut off (Bier et al. 2004). The ACE has also been shown to discriminate, albeit imperfectly, between elderly patients with dementia and those with affective disorders (Dudas et al. 2005a) and when administered in conjunction with measures of episodic memory, appears to be of utility in helping to identify aMCI sufferers who will convert to AD in the short term (Ahmed et al. 2008b).

2.2.2 Episodic Memory

2.2.2.1 CANTAB Paired Associate Learning Test (PAL; (Morris et al. 1987))

The Paired Associate Learning (PAL) subtest from the Cambridge Neuropsychological Test Automated Battery (CANTAB; (Morris et al. 1987)) is a computerised measure of visuospatial learning requiring participants to learn the locations of an increasing number (i.e. 1, 2, 3, 6, and then 8) of abstract patterns. The score of interest was the number of pattern-position errors made at the 6-pattern level as this has been previously shown to be sensitive to episodic memory impairment in MCI (Ahmed et al. 2008b;Alladi et al. 2006) as well as Questionable Dementia patients but not to depression (Swainson et al. 2001). The total number of errors made at the 6 pattern level is also reportedly highly specific to memory impairment arising in from AD (O'Connell et al. 2004), although perhaps less sensitive (sensitivity = 68%) to early AD than initially reported.

The PAL subtest is language free, reducing the potential for cultural bias. Trials are graded in terms of their level of difficulty to ensure that instructions have been comprehended and to avoid floor and ceiling effects. The computerised administration assists in ensuring that the test is given in a standardized manner. Normative data are available for 3000 healthy volunteers, banded according to age, gender and pre-morbid IQ.

Acceptable to high levels of concurrent validity (0.39 – 0.67) and test-retest reliability (0.64 – 0.88) have been reported for the PAL subtest of the CANTAB battery (Fowler et al. 1995). The PAL subtest is reportedly sensitive to small changes in episodic memory function across even relatively brief i.e. 6-12 month intervals (Fowler et al. 1997). The construct validity of

the PAL subtest as a measure of medial temporal lobe integrity is well established by way of both animal lesion (Fray et al. 1996) & fMRI studies (Gould et al. 2005).

2.2.2.2 Rey Complex Figure Test (RCFT; (Rey 1941))

This measure was designed to assess both visuospatial skill and visual memory (Rey 1941). In this task, participants are given a complex figure and are asked to make a copy of it without time restriction. Immediately after presenting the figure and again following a 30-minute delay, participants are required to reproduce the figure from memory. Three measures of interest were obtained including (each with a total possible score of 36 points); the copy score, the immediate recall score and the delayed recall score out of a possible total of 36 points. Healthy control subjects generally obtain similar scores on the immediate and delayed components of the RCFT, reflecting an ability to retain the bulk of new information that has been encoded over the short term. By contrast, poor performance on the recall components of the RCFT has been observed in AD patients (Lezak 2004), in keeping with the encoding and consolidation deficits that are known to characterise this patient group.

The effect of age appears to be minimal on the copy component (Mitrushina et al. 1999), however, a decline in recall performance is seen in association with both advancing age and fewer years of education. High inter-scorer reliability values of between r = 0.91 to 0.98 (Lezak 2004) and r = 0.80 to 0.99 (Mitrushina et al. 1999) have been reported for all components of this task. Test-retest reliability coefficients are somewhat lower, ranging between r = 0.56 to 0.68 for the copy and from r= 0.57 to 0.77 for the 3-minute delayed recall component, when administered thrice yearly (Mitrushina et al. 1999).

2.2.2.3 Hopkins Verbal Learning Test – Revised (HVLT-R; (Brandt 1991))

All groups completed the Hopkins Verbal Learning Test – Revised (HVLT-R; (Brandt 1991) a test of verbal episodic recognition and recall memory and new learning ability. Participants are required to recall as many words as possible immediately following presentation of a 12-item word list across three consecutive learning trials. 3 categories, i.e. ‘Animals’, ‘Precious Stones’ & ‘Dwellings’, with four words belonging to each, make up the

12-item list. Measures include; total number of words recalled across three registration trials (max=36), total number of words recalled following a 30-minute delay (max=12) and a discrimination index score representing a participant's ability to discriminate between old and new list items. The total number of words recalled immediately following list presentation provides a measure of new learning ability, while the number of words recalled after a short delay serves as a measure of delayed verbal recall ability.

The HVLT-R has been shown to be sensitive and specific to the early stages of AD (sensitivity = 98%; spec=95%) as well as to MCI (sensitivity = 78%; spec = 80%; (De Jager et al. 2003). The absence of ceiling effects in a 70-88 year old age group (Lezak 2004), high sensitivity and specificity values in relation to dementia, together with a lack of need to adjust for education level (Vanderploeg et al. 2000), make it of potential utility as a dementia screening measure (Hogervorst et al. 2002). Validity of the total recall and recognition discrimination index has been established by way of significant relationships between these and corresponding indices from other well established episodic memory measures (Lezak 2004).

Significant learning effects (Z = 5.80) for the total number of words recalled across the three trials of version 1 of the HVLT have been demonstrated across 2-3 year periods in healthy elderly controls (Schrijnemaekers et al. 2006) but not in early AD or MCI sufferers. Significant negative associations between age and male gender but not education and HVLT total trials recall for Form 1 have been documented in a large sample of older adults (Vanderploeg et al. 2000). Statistically significant 9 month stability coefficients of a similar magnitude to those reported for other clinical episodic memory measures have been documented for total recall (r = 0.50), true (r = 0.66) and false positives (r = 0.42) in a small group (n=45) of healthy elderly adults (Rasmusson et al. 1995).

2.2.3 Semantic Memory and Language

Participants completed the Graded Naming Test (GNT; (McKenna and Warrington 1980)), the Graded Faces Test (GFT; personal communication, 2004), the Boston Naming Test (BNT; (Kaplan et al. 1983)), the Edinburgh Exemplar Naming Test (EENT: JAL Author)

and the Category Fluency Test with the categories 'Animals', 'Fruits' & 'Vegetables'(Lezak 2004).

2.2.3.1 Graded Naming Test (GNT; (McKenna and Warrington 1980))

The GNT requires participants to name 30 line drawings of increasing difficulty, with early items named by control subjects without difficulty. There is a mixture of animate and inanimate objects. Items comprising the latter part of the test are less familiar and less frequently encountered such that many normal controls fail to name these items. Unlike the BNT, many items might be considered biased toward the British Culture (i.e. sporran, mitre, tassel, latch and bellows). Of the 100 normal control subjects (age range 18-76) comprising the normative group for the initial validity study, scores ranged from 20-100% (McKenna & Warrington, 1980).

Pre-morbid IQ has been shown to correlate significantly with GNT total score (Bird et al. 2004) however no effect of age was observed in this sample of 188, 40-70 year old healthy adults and only a minimal age effect was noted in a more recent standardisation sample comprising 305 adults between the ages of 18-77 years (Warrington 1997). The authors of the former study simultaneously documented good test-retest reliability ($r = 0.92$), significant yet moderate practice effects (in the order of one additional point) and a small reliable change index score of 5 points. The sensitivity of the GNT to gradual changes in performance within a neurological sample across a two year period was also established (Bird and Cipolotti 2007).

2.2.3.2 Boston Naming Test (BNT; (Kaplan et al. 1983))

The BNT requires participants to name 60 line drawings of increasing difficulty. A 20 second naming interval is provided, following which a semantic cue is provided, along with another 20 second naming interval (i.e. for harmonica – it is a musical instrument) in cases where no response has been made. The black and white line drawings comprise a mixture of animate and inanimate objects. The measure was administered in keeping with standardised clinical practice where participants were first asked to name item 30 and continued with subsequent items in their correct numerical order in cases where the initial 8 responses provided were correct. Where errors are recorded prior to reaching item 37, the examiner

administers items in a backward direction from 29, up until 8 consecutive correct responses have been recorded. Following this, the remainder of the test items post item 30), are administered. The test was terminated following the recording of 8 incorrect consecutive item responses.

Mitrushina and Satz (Lezak 2004) report an absence of practice effects on the BNT, at one year intervals. High response consistency (80%) has been documented in AD patients at 6 month retest intervals (Spreen and Strauss 1998). Whilst a significant positive relationship between education and BNT performance is known to exist, the association with gender is at best inconsistent (Lezak 2004). Minor age related declines in BNT performance are thought to appear relatively late (i.e. from the age of 70 onwards, perhaps later i.e. 80+ in more highly educated elderly) alongside increased variability among the elderly normal population (Lezak 2004).

2.2.3.3 Edinburgh Exemplar Naming Test (EENT; JAL)

The EENT was developed by the author (JAL) in an effort to improve the sensitivity of existing confrontation naming measures to early semantic memory failure. In this test, the participant is required to name 50 line drawings of low frequency, animate objects with sizable feature overlap. One point is received for every correct animal name provided. Participants obtain a score out of a total possible 50 points. It is assumed that animate members of a single exemplar category sharing appearance and activity based features will prove more difficult for early and pre-clinical AD patients to name than items belonging to unique categories, without shared features. This predicted pattern of test failure would fit well with what is understood of the graded manner in which the semantic system fails in the temporal lobe variant of the Frontotemporal lobe dementias, where degradation in detail pertaining to the different exemplars belonging to a specific category is observed prior to the point at which the patient demonstrates a complete loss of knowledge for the category as a whole. The psychometric properties of this test have not been established.

2.2.3.4 Graded Faces Test (GFT; personal communication, 2004)

The GFT requires participants to name a series of 30 black and white photographs of famous faces. The faces comprise a combination of famous politicians, statesperson, actors,

musicians or athletes. Fifteen of these were recently famous people and 15 were non-recent celebrities. Thirty faces were selected from a database of 250 famous faces spanning the second half of the 20th century. The items were graded in difficulty to match the items from the GNT (i.e. In the GNT 'buoy' was named by 97% of controls; In the GFT 'Adolf Hitler' was named by 97% of controls). The age of the items (i.e. recent versus non-recent) was balanced throughout the levels of difficulty. One point is given for every correct full name provided such that participants obtain a score out of a total possible 30 points (Thompson et al. 2002).

There is no published data pertaining to the relationship between age, IQ, education or gender or test re-test reliability and performance on the GFT. Preliminary normative data comprising 22 healthy elderly (mean age = 66 years 11 months SD 7 years 12 months) community dwelling, highly educated UK residents, were obtained by way of personal correspondence with Dr Kim Graham, a co-developer of the GFT. The mean total naming score for this sample was 22.09 (standard deviation = 3.28). The sensitivity of face naming tasks to early AD and MCI has been well documented within the test developers research group (Ahmed et al. 2008a;Clague et al. 2005;Dudas et al. 2005b;Thompson et al. 2002).

2.2.3.5 Category Fluency (Animals + Fruits + Vegetables; (Lezak 2004))

Participants were also asked to complete a Category Fluency Task, requiring them to produce as many animals as they could in a one minute time period. The procedure was then repeated for fruits and finally for vegetables. Two scores were derived; one representing the total number of animals generated in one minute (Animals) and another representing the total combined number of animals, fruits and vegetables produced across their respective one minute time intervals (Total Category Fluency).

Moderate one month test re-test reliabilities (r = 0.56) & practice effects in the order of just over 1 point together with large reliable change indices (lower = -7.6; upper = 10.5) and an absence of correlation with age or IQ have been reported in association with the animal fluency task (Bird et al. 2004).

2.2.4 Attention, Executive Function & Processing Speed

2.2.4.1 Trail Making Test (TMT A and B; (Reitan 1985))

All participant groups completed the Trail Making Tests, Parts A and B (TMT A & B; (Reitan 1985). This test provides a measure of scanning, visuomotor tracking, divided attention and cognitive flexibility (Lezak 2004). Participants were required to join numbered circles in ascending order (Part A) and numbers and letters in ascending alternating sequence (Part B) at pace and the time to completion is recorded. The time taken for participants to complete TMT A represents a measure of psychomotor processing speed, while TMT B also adds a divided attention component. By subtracting the time to completion for TMT A of this test from TMT B, a measure of the 'executive' functioning component, or 'attentional demands' can be acquired independently of processing speed.

According to Lezak (Lezak 2004), there is a positive correlation with age and TMT performance with performance times increasing significantly with each succeeding decade. By contrast, performance on Part B of the TMT correlates negatively with years of education. Reported reliability coefficients vary with most above r = 0.6 and several in the r= 0.9 and more in the r= 0.8 range. Practice effects are usually seen on both parts of the TMT with effects on Part A more likely to reach significance as a result of the greater variability in TMTB performances. However, an absence of practice effects has been documented across longer i.e. one year test re-test intervals in one study. Depression has been found to have a slowing effect on TMTB that interacts with age, such that elderly depressed patients require a disproportionately greater amount of time to complete this than emotionally stable elderly subjects or depressed younger ones. The normative distribution for TMT scores is known to be positively skewed such that cut off (as opposed to graded) scores are considered most appropriate for use within a clinical context (Lezak 2004).

2.2.4.2 Dual Performance Task (Della Sala, personal communication, 2004; (Della Sala 2005)

All participants completed the revised dual performance task which was generously provided to us by Della Sala and colleagues (Della Sala, personal communication, 2004; (Della Sala 2005). This pencil and paper test of divided attention was created to tap functions associated with the central executive component of working memory. It consists of 2 tasks (a digit span task and a visuospatial tracking task) that are each performed on their own before being

performed concurrently. Firstly, participants' digit span was determined. This involved repeating strings of digits read by the test administrator at a rate of approximately 2 per second. Initially, 2-digit strings were presented and these increased one digit at a time on condition that the participant correctly recited 5 of 6 examples of each length. When the participant failed to recite 2 or more strings of the same span, the digit span for that person was considered to be the previous length. No time limit was used at this stage. Having determined the participants' individual digit span, participants had 90 seconds to recite as many digit strings (fixed at the individual participants' digit span) as possible. Responses were recorded as correct for each digit recited in the correct order.

Following this, participants completed the tracking task. This involved participants tracing a line through circles on an A3 sized sheet of paper in the experimental trail, with a short practice session prior to this. Participants had 90 seconds for this trial, and the number of circles reached during this time was recorded. The final trial was the concurrent dual task. Here participants had 90 seconds to simultaneously recite digit strings fixed at their digit span as well as carrying out the visuo-spatial tracking task. In order to take into account the various strategies one may adopt in performing the two tasks simultaneously, an overall decrement score was calculated using the following formula:

$$\mu = [1-(P_m + P_t/2)] * 100$$

where μ is the combined dual task score, P_m is the proportional loss in span performance between single (P_s) and dual task (P_d) conditions, $[(P_s - P_d]/ P_s]$ while P_t is the equivalent proportional tracking score. Thus a score of 100 would represent no dual task decrement and lower scores reflect greater dual task decrements.

Correlations between dual task performance, gender and education are reportedly insignificant and re-test reliability values moderate at best (r=0.49; (Rabbitt 1997).

2.2.4.3 Controlled Oral Word Association Test (COWAT; (Spreen and Strauss 1998))

All participants completed the COWAT using the letters 'F', 'A' & 'S'. The participants were required to generate as many words as possible beginning with each of these letters, within separate one minute time frames. Further instruction was given not to include names of people or places or proper nouns in general. Participants were also required to produce different words rather than use the same words with different endings. The number of words produced for each letter was added to give a 'Total Verbal Fluency score'.

A similar procedure was also undertaken using the letter 'P', within the context of the ACE. Here, the number of words generated beginning with the letter 'P' were summed to give a 'P words Verbal Fluency score'.

There is variability in the reported effects of age on COWAT scores in the elderly, with some studies revealing a significant negative relationship and others no relationship (Lezak 2004;Spreen and Strauss 1998). Education may be a moderating factor as no significant effect of age was noted in a group of highly educated elderly from 75 years onwards whilst a significant decline was noted in a less educated subgroup from the age band 50-54 years (Lezak 2004). The total number of words produced on the COWAT has been shown to correlate with performance on the NART ($r = 0.67$) whereas gender appears to have no discernable effect on COWAT scores. A one year test re-test reliability of $r = 0.71$ has been reported for a healthy elderly sample and inter-scorer reliability is reportedly near perfect. Practice effects of one point together with a test re-test reliability of $r = 0.65$ has been documented in a sample of adult epileptic patients, retested following 8 months. Slightly higher re-test reliabilities ($r = 0.76 – 0.87$) were reported in a sample of HIV positive adults tested three times across an 18 month period (Spreen and Strauss 1998).

2.3 Functional and Mood Scales

2.3.1 Measure of Everyday Functioning (Farias et al. 2006)

The Measure of Everyday Functioning is an informant based questionnaire comprising a total of 30 statements for which the primary carer is required to rate each statement by comparing current abilities to those observed 10 years ago. The informant ascribes a score

of between 1 and 4 where 1 represents an improvement or no change in the designated ability and 4 indicates that the patient is consistently much worse. A score of 0 is marked when the skill in question does not apply to, or form part of the patients' everyday repertoire of activities. Statements are grouped in accordance with their reliance on the integrity of different cognitive abilities including memory, language, planning, organisation and divided attention.

Functional items were selected from a range of existing scales on the basis of their ability to assess high-level everyday skills. A rating scale was applied (as described above) in place of a dichotomous 'dependent' versus 'independent' system of scoring to facilitate measurement of more subtle or mild changes in function that fall short of leading to complete dependence on others. A reference point of 10 years prior was chosen as it was thought to provide enough of a time lapse that no deterioration in function would likely have been present at that time (Farias et al. 2006).

For the purpose of this study, both raw total scores and total score as a proportion of the items of personal relevance (i.e. after omitting those items deemed no to apply) were calculated for every aMCI patient at his/her final follow-up assessment, where higher scores signify greater levels of impairment of higher-level everyday activities.

Previous findings from the validation study indicate that MCI participants exhibit significantly greater levels of impairment than healthy age matched controls on 73% of items on the everyday memory scale, 61% of items on the everyday planning and 55% on the everyday organisational items, 28% of the everyday language and 22% of the everyday divided attention items. The dementia group was significantly more impaired than both the MCI and healthy elderly control group on all 74 items comprising the scale (Farias et al. 2006)

2.3.2 Personal Self Maintenance Scale (PSMS; (Lawton and Brody 1969))

The PSMS is a six item scale, rating self-care ability in the domains of feeding, grooming, bathing, toileting, personal hygiene and dressing. Primary care givers are required to

endorse the appropriate rating (1-5) for each area of personal care, where increasing ratings reflect poorer self maintenance and independence levels. As all items require a rating of a least 1, only a single total rating score was derived for this measure, with greater overall scores reflecting lower personal self maintenance abilities.

The scale's validity has been established by way of correlations with several other established measures of functional health, cognitive status, behaviour and adjustment rating scales within a group of 180 care home applicants. Reproducibility coefficients of 0.96 are reported for the PSMS alongside inter-rater reliabilities of between 0.87 and 0.91 (Lawton and Brody 1969).

2.3.3 Independent Activities of Daily Living Scale (IADLS; (Lawton and Brody 1969))

The IADL scale is an 8 item scale assessing the patient's ability to perform the following tasks; using the telephone, preparing food, shopping, laundering, completing household tasks, responsibility for medications, management of financial matters and mode of transportation. The patient's primary care giver is required to denote, on a scale of between 1 and 3, 4 or 5, the level at which the patient is able to engage in, or carry out activities in each of these areas. Higher scores signify a greater degree of dependence on others and a score of 0 implies that the activity in question is not applicable in that it would not previously have been performed by the patient.

As the total score is somewhat dependent on the applicability of the activities in question to each individual patient, two scores were derived, 1) a total score, where higher scores are broadly indicative of greater levels of functional dependence and 2) a proportionate score, where the total score is expressed as a function of the maximum number of available points (i.e. following omission of any points allocated to any activity that has been endorsed as 'not applicable').

The scale's validity has been established by way of correlations with several other established measures of functional health, cognitive status, behaviour and adjustment rating scales, within a group of 180 care home applicants. Extensive testing of the reliability of the

IADL has not been conducted, although high correlations between independent ratings for 12 elderly nursing home applicants, based on information obtained within a single interview, have been documented (Lawton and Brody 1969).

2.3.4 Geriatric Depression Scale (GDS; (Yesavage et al. 1983))

The GDS is a 30 item 'yes/no' answer scale that was constructed for the purposes of brief screening for depression amongst the elderly, with less emphasis on somatic symptoms, age appropriate questions and a simplified method of response. Of the 30 items, 20 indicate the presence of depression when answered positively, while 10 others indicate depression when answered negatively (Yesavage et al. 1983). The number of endorsed items is summed, with higher scores, particularly those greater than 10, being associated with depression or the possibility thereof (Brink et al. 1982). The former authors reported sensitivity and specificity values to depression of 84% and 95%, respectively applying a cut off score of 11/30. In the original validation study, normal elderly controls obtained a mean total score of 5.75 (SD = 4.34) as compared to significantly higher scores of mildly (15.05; SD = 6.50) and severely depressed (22.85; SD = 5.07) elderly outpatients (Yesavage et al. 1983). High levels of validity and one week test re-test reliability (r = 0.85) were also reported.

2.4 Outcome Variables

In recognition of the variety of possible longterm outcomes seen in association with aMCI (Ahmed et al. 2008b;Lee et al. 2006;Schmidtke and Hermeneit 2007), at the final follow-up assessment, MCI participants were assigned to 1 of 4 groups in accordance with whether their cognitive deficits had remained stable - stable MCI; their psychometric performance was normal - normal; They had declined psychometrically over the course of their participation in the study but did not meet DSM-IV criteria for dementia – decline; or they had declined sufficiently on a cognitive level and exhibited associated areas of functional impairment to fulfil DSM-IV or other published non-Alzheimer consensus criteria for dementia - dementia. Assignment to each of these four groups was carried out in accordance with the criterion outlined below.

2.4.1 Stable MCI

Stable MCI participants were those exhibiting evidence of impaired episodic memory function at baseline and subjective memory complaints, with or without deficits in additional non-

episodic memory domains, who continue to perform at least 1.5 SD below norms on two or more episodic memory measure at final follow-up but do not show evidence of significant cognitive decline in non-memory domains (as defined by a deterioration in performance on two or more executive or semantic memory measures of a magnitude found in less than 2.5% of an age/IQ matched sample over an average 3.17 year period).

Criterion

1.5 SD or more below control mean on two or more neuropsychological measures at final assessment

AND

No evidence of significant declines in non-memory domains (as defined by a decline of a magnitude that would be expected to occur in less that 2.5% of a normative sample over the course of the study on two or more tests within a given cognitive domain).

2.4.2 Normal

Normal participants were those performing within 1.5SD of the control mean on at least 7 out of 8 episodic memory measures at final follow-up.

Criterion

1.5 SD or more below control mean on a maximum of one neuropsychological measure at final assessment

2.4.3 Declining MCI

Declining MCI participants were performing at least 1.5 SD below age norms at baseline on one or more measures of episodic memory function or at least 1 SD below age norms at BL

on two or more measures of episodic memory function (+ or – other domains of cognitive impairment) and exhibiting decline of a magnitude that would be expected to occur in less that 2.5% of a normative sample over the course of the study on either two measures of semantic memory functioning or two measures of executive ability. Declining MCI participants did not meet criteria for dementia on account of their sound functional abilities.

Criterion

1.5 SD or more below control mean on two or more neuropsychological measures at final assessment.

AND

Evidence of significant declines in a non-memory domain(s), as defined by a decline of a magnitude that would be expected to occur in less that 2.5% of a normative sample over the course of the study on two or more measures assessing a specific non-memory domain i.e. semantic memory or executive function.

AND

No evidence of significant functional decline (as defined by 3 or more ratings of 3 or 4 in 2 or more non-memory domains on the MCI ADL scale).

OR

No evidence of decline in two or more areas of the ADL scale.

The different thresholds for varying items of the ADL scale were applied in an attempt to capture the point at which the level of performance for each item was suggestive of functional impairment. For example, for item C, relating to food preparation, it is plausible that a married 70-year-old man, without functional impairment, does not 'plan, prepare and serve adequate meals independently (1)', or indeed 'prepare adequate meals if supplied with ingredients (2)' whereas an inability to maintain an adequate diet, would be suggestive of some significant functional difficulty.

Memory complaints were universal amongst the MCI cohort at baseline and DSM-IV clinical criteria for dementia requires evidence of additional non-memory domain(s) of cognitive impairment that cause significant impairment in social and/or occupational

functioning. As such, patients whose carer's were not (1) rating several aspects of their non-memory functioning as 'consistently a little (3) or much worse (4)' at the study endpoint, or (2) endorsing clear evidence of decline in more than one activity listed in the Instrumental Activities of Daily Living Scale (Lawton and Brody 1969), were considered insufficiently functionally impaired to fulfil DSM-IV criteria for dementia.

2.4.4 Dementia

Dementia was diagnosed in accordance with DSM-IV criteria for AD where the development of multiple cognitive deficits manifested by both memory impairment together with one or more of language disturbance, apraxia, agnosia or executive functioning and criteria for continuing cognitive decline were defined by the following criteria.

Criterion

1.5 SD or more below control mean on one or more episodic memory measures (ACE immediate recall, ACE delay, PAL 6 box errors, HVLT total trials, HVLT delayed recall, HVLT Discrimination Index (DI) score, RCFT immediate recall, RCFT delay).

OR

Meets published clinical criteria for one of the non-AD or vascular forms of dementia.

AND

Evidence of significant declines in non-memory domains (as defined by a decline of a magnitude that would be expected to occur in less that 2.5% of a normative sample over the course of the study on two or more measures assessing either semantic memory or executive function.

That cognitive deficits cause significant impairment in social or occupational functioning and represent a significant decline from a previous level of functioning was defined by the either of the following;-

Criterion

Evidence of significant functional decline (as defined by 3 or more ratings of 3 or 4 in 2 or more non-memory domains on the MCI ADL scale).

OR

Evidence of decline in two or more areas of the ADL scale (Lawton and Brody 1969), rated in accordance with criteria previously described (see Appendix 8.4.2 for copy of the Instrumental Activities of Daily Living Scale)

All MCI participants had undergone one or more prior consultation(s) with their general practitioner as well as either a Consultant Old Age Psychiatrist's or a Consultant Geriatrician, prior to being referred for neuropsychological assessment. In addition, 22/46 aMCI participants underwent some form of neuroimaging facilitating the screening out of other central nervous systems and systemic conditions that can cause progressive deficits in memory and cognition. All participants who were categorised as having developed dementia by the study endpoint fulfilled the criterion below.

Criterion

1.5 SD or more below control mean on two or more episodic memory measures (ACE immediate recall, ACE delay, PAL6 box errors, HVLT total trials, HVLT delayed recall, HVLT DI, RCFT immediate recall, RCFT delay).

AND/OR

1.5 SD or more below control mean on two or more non-episodic memory measures (EENT, GNT, BNT, GFT, Category fluency total, FAS total, TMTA, TMTB, Dual Performance Combined score, RCFT copy).

AND

Displaying evidence of significant declines in non-memory domains (as defined by a decline of a magnitude that would be expected to occur in less that 2.5% of a normative sample over the course of the study on two or more tests of a cognitive domain other than that of (or in conjunction with) memory.

AND

Displaying evidence of significant functional decline (as defined by 3 or more ratings of 3 or 4 in 2 or more non-memory domains on the MCI ADL scale).

OR

Evidence of decline in two or more areas of the IADL scale (Lawton and Brody 1969).

Diagnoses of other non-AD forms of dementia (Vascular dementia, VD; frontal variant Frontotemporal dementia, fvFTLD; Semantic dementia, SD; Primary progressive aphasia, PPA; Dementia with Lewy Bodies, DwLB) were made in accordance with established criteria (2002;McKeith et al. 1999;Mesulam 2001;Neary et al. 1998) by way of reference to the existing neuropsychological data within the context of a clinical interview with the patient and his/her carer at the final follow-up assessment, or within the context of a clinical neuropsychological consultation.

2.5 Additional Medical Information

We also recorded the number of MCI patients who were receiving anticholinesterase and/or antidepressant drug therapy at their final assessments (see Table 8.3 Appendix B). The level of service use at the time of the final follow-up assessment was recorded for each of the 46 MCI participants, both in terms of the type of professional contact(s) they were receiving (i.e. GP, Consultant Psychiatry, Clinical Psychology, Community Psychiatric Nurse, Social Services or none) and the frequency of consultations over the past 12 month period. Note was also made of any co-morbid medical conditions that were documented in the medical notes. Where a clinical diagnosis of dementia had been made post study entry, this was also recorded along with type, where this had been specified. The above details are tabulated in section 8.2 of the Appendix.

3. The Prevalence and Neuropsychological Characterisation of aMCI

(Published in part in British Journal of Psychiatry. 2008 Jan; 192:59-64.)

3.1 The Prevalence of aMCI in Specialist Memory Clinic Settings

3.1.1 Abstract

Background

There is current interest in exploring the different subtypes of mild cognitive impairment (MCI) in terms of their epidemiology.

Aims

To examine the frequency of MCI subtypes presenting to a memory clinic.

Method

Consecutive tertiary referrals (n=187) were psychiatrically evaluated; 45 patients met criteria for amnestic mild cognitive impairment (aMCI).

Results

Patients who fulfil Petersen's criteria for MCI represent one fifth of all referrals made to our specialist memory clinic.

Conclusions

Amnestic MCI is an important diagnosis in secondary and tertiary memory clinics.

3.1.2 Introduction

The practical utility of screening for MCI within the general population has been questioned, on grounds that a significant portion of those who are identified as suffering with MCI either do not wish to pursue any further investigation, or have an identifiable medical cause for their cognitive impairment (Jicha 2008). By contrast, those patient's who have been referred to specialist memory clinics and found to fulfil criteria for MCI, have actively sought consultation, are more likely to have undergone relevant physical screening to rule out any medical cause for their symptoms, and are arguably of clinical importance in this regard.

Whilst a number of studies have reported MCI prevalence estimates for the general population (Blossom et al. 2007;Busse et al. 2003;Panza et al. 2005;Pioggiosi et al. 2006), there is little information regarding the frequency with which MCI is encountered at a specialist clinical level (i.e. within a memory clinic setting). This is important to know, as there is evidence to suggest that both assessment (Lonie et al. 2008) and treatment (Akhtar et al. 2006;National Collaborating Centre for Mental Health 2006) approaches for MCI may differ from those that are recommended in established dementia cases.

In view of the ageing population, the availability of memory treatment agents, and the current clinical and research emphasis on early diagnosis, one might expect that MCI referrals comprise a significant proportion of overall referrals received within specialist memory clinic settings. The results of existing studies suggest this may be the case. Lehrner et al., (Lehrner et al. 2005) reported that approximately one-fifth of the patients seeking help in their outpatient memory clinic were identified as having aMCI. A similar proportion of MCI type referrals (23%) were also received by the Cambridge Memory Clinic across a 21 month period (Alladi et al. 2006). A lower proportion of MCI referrals (12.7%) have, however, been reported more recently, in connection with a newly established UK memory clinic (Popoola et al. 2008).

As such, whilst there is reason to suspect that a significant portion of the patients presenting to specialist memory clinics might fulfil criteria for MCI, there presently exists a small amount of conflicting data informing the prevalence of MCI within memory clinic settings.

As a prelude to cross-sectional and longitudinal neuropsychological analyses of our aMCI cohort, we sought to determine the proportion of referrals made to The Edinburgh Neuropsychological Assessment Service for Older People fulfilling criteria for aMCI. On the basis of figures reported by two other well established neurology lead memory clinics, we hypothesised that patient's fulfilling aMCI criteria (Petersen et al. 1999;Petersen 2004;Petersen 2005a) would comprise a significant (i.e. >20%) proportion of consecutive referrals received by our clinic across an 18 month time interval.

3.1.3 Method

We retrospectively analysed 187 consecutive referrals to the Edinburgh Neuropsychological Assessment Service for Older People between the months of September 2004 and April 2006. Referrals were received at a tertiary level, stemming from consultants in older age psychiatry, geriatric medicine and neurology. All of these patients had undergone comprehensive psychiatric evaluation, relevant medical screening (including a standard battery of screening blood tests) and neuroimaging (computed tomography and/or magnetic resonance imaging or single-photon emission computed tomography) prior to being referred to our service. All but three patients were over the age of 50 years.

The original criteria for MCI set out by Petersen et al (Petersen et al. 1999) require that a person must present (1) with a memory complaint, (2) show evidence of objective memory decline in relation to age and education, (3) demonstrate preservation of other areas of cognitive function and (4) activities of daily life and (5) not fulfil criteria for dementia. Because it has since become apparent that not all persons who demonstrate cognitive impairment short of dementia have a 'memory' complaint, we utilised the recently expanded criteria that include persons with non-memory complaints (single domain non-memory MCI) as well as those exhibiting multiple domains of cognitive impairment who nonetheless fail to fulfil criteria for dementia (multiple domains slightly impaired) (Petersen, 2005;Petersen, 2004). The Mini-mental State Examination (MMSE; Folstein et al, 1975) and Addenbrookes Cognitive Examination (ACE; Mathuranath et al, 2000) were administered as a means of establishing the patient's general level of cognitive functioning. Level of everyday functioning was examined by way of the Clinical Dementia Rating Scale (Morris,1993)

within the context of a clinical interview with the patient and his/her primary carer, where available.

A total of 112 patients fulfilled either one or more of the following exclusion criteria and were therefore excluded from the analyses: dementia (MMSE < 24/30 or ACE <80/100 and DSM-IV, depression (as assessed either by way of formal psychiatric consultation or, in a small proportion of cases, by a score of greater than 10 on the Geriatric Depression Scale (GDS; Yesavage et al, 1982) or clinical assessment by the author (JAL), or one or more medical or psychiatric conditions that could conceivably account for the patient's cognitive impairment (i.e. head injury, schizophrenia, evidence of stroke or tumour on neuroimaging, alcoholism, epilepsy, cranial radiotherapy)

Of the remaining 75 patients, 15 showed cognitive impairments outside the domain of episodic memory, 15 returned a 'normal' cognitive profile and 45 showed memory function impaired for age (with or without additional areas of cognitive impairment).

3.1.3.1 Neuropsychological Assessment Measures

Details of the neuropsychological assessment battery are provided in section 2.2 of Chapter 2 Materials and Methods.

3.1.3.2 Comparison with other memory clinics

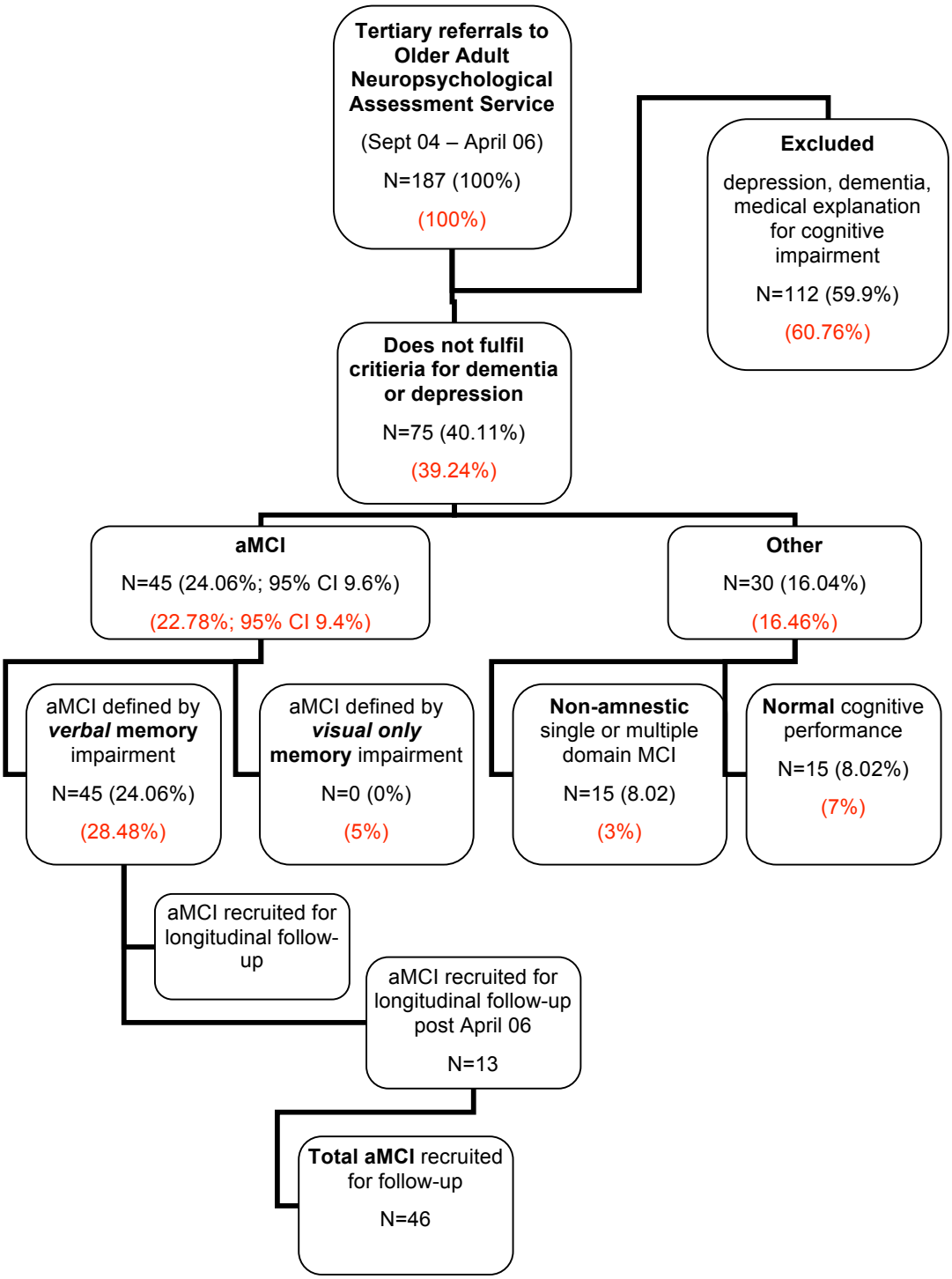
We searched the literature for studies employing neuropsychological test batteries similar to ours to examine the comparability of our sample with other published data.

3.1.4 Results

3.1.4.1 Results of literature search

We identified one other study reporting consecutive referrals to a memory clinic using similar diagnostic criteria and assessment measures (Alladi et al, 2006).

Figure 3.1 Flowchart of aMCI referrals to the Edinburgh Older Adult Neuropsychological Assessment Service (September 2004 – April 2006)



3.1.4.2 Comparability of aMCI Referral Patterns

A striking similarity in referral patterns was observed between our Neuropsychological Assessment Service for Older Adults and data reported from the Cambridge Memory Clinic (Alladi et al. 2006). Data from the latter study are presented below our figures in percentage format in red (Figure 3.1), for comparative purposes. When the 150 pre-excluded referrals from the Cambridge Memory Clinic were accounted for, just over half (i.e. 60%) of referrals from both centres were excluded due to an established dementia or depressive disorder or one or more medical condition(s) that could account for the patient's cognitive impairment ($chi^2 = 0.015$; $p= 0.90$). Close to 40% of referrals from both centres fell within the non-demented and non-depressed category ($chi^2 = 0.000$; $p = 0.995$). Just over half of these patients in both centres met Petersen's (1999 & 2004) criteria for aMCI ($chi^2 = 0.46$; $p = 0.50$) representing close to one fifth of overall referrals from both centres. Of the remaining 40% of non-demented, non-depressed patients, half demonstrated cognitive deficits of a non-amnestic variety (in one or more domains), and half returned 'normal' cognitive profiles.

Although referral patterns for aMCI were similar across the two centres, there was a greater proportion of non-amnestic MCI patients and fewer visual only aMCI patients in our sample ($chi^2 = 13.23$ (3 df); $p = 0.004$; Fisher-Freeman-Halton exact test $p =0.003$).

3.1.5 Discussion

In this study we have shown that non-depressed, non-demented persons who fulfil Petersen's (1999 & 2004) criteria for aMCI, make up a significant proportion of referrals to our Old Age Clinical Neuropsychology service. Roughly one quarter of all referrals received across an 18-month period met Petersen's (1999 & 2004) criteria for aMCI. This is an almost identical proportion of patients to that reported in a recent study by the Cambridge Memory Clinic (22.78%) (Alladi et al, 2006) and a very similar figure to the 20% of patients seeking help in Lehrner et al's (2005) clinic who met criteria for aMCI.

It is conceivable that the considerably lower MCI referral rate reported by the Cork University Hospital Memory Clinic was a reflection of its recent establishment, or the fact that this clinic, in contrast to the others, was Psychiatry lead. Surprisingly, the level at which referrals are received (i.e. whether primary, as in the Cambridge Memory Clinic or tertiary, as in our clinic) appears not to influence the proportion of MCI-type referrals that are

received. Thus, although by no means universally adopted, it appears that both the concept and criteria for aMCI are applicable and, indeed, a necessary adjunct to clinical practice.

It is unusual for MCI patients to present with predominant psychiatric or neurological complaints, and pharmaceutical options are presently unavailable to MCI sufferers on the National Health Service (NHS). By contrast, there are a number of well established cognitive rehabilitative techniques that have been validated for use in MCI (Akhtar et al. 2006); Kapur and Wilson, personal communication, 2008). Furthermore, traditional dementia screening measures that are commonly utilised as a means of establishing a patient's overall level of cognitive functioning are insensitive to cognitive deficits in MCI (Lonie et al. 2008) and influenced by IQ (Starr and Lonie 2007). MCI patients may perform within normal age limits on even the most comprehensive of these (Alladi et al. 2006;Lonie et al. 2008).

The above observations, together with the frequency with which MCI is encountered at specialist levels, raises the question of whether a different or wider mix of skill base may be necessary, within specialist memory clinics, to meet the more comprehensive psychometric assessment and non-pharmaceutical treatment needs of this patient group.

3.1.6 Conclusion

MCI is an important diagnosis within specialist secondary and tertiary referral based memory clinic settings. The large proportion of referrals of this nature received by our own and other memory clinics highlights a need for evidence-based guidance as to how these patients might best be managed clinically and more specifically, neuropsychologically.

3.2 Neuropsychological Characterisation and Classification of aMCI

(Published in part in British Journal of Psychiatry. 2008 Jan; 192:59-64.)

3.2.1 Abstract

Background

There is current interest in exploring the different subtypes of mild cognitive impairment (MCI), in terms of their cognitive profile.

Aims

To document detailed neuropsychological profiles of patients with the amnestic subtype.

Method

Consecutive tertiary referrals (n=187) were psychiatrically evaluated. A group of 46 patients with aMCI as well as 24 healthy controls took part in a thorough neuropsychological examination.

Results

Of the 46 aMCI patients who were examined in greater neuropsychological detail, 36 were performing 1 SD or more below the control mean on 2 or more episodic memory measures and 39 were performing 1.5 SD or more below the control mean on 1 or more of these measures. Four returned a psychometrically 'normal' profile. A majority, i.e. 32-34 were exhibiting evidence of additional non-memory impairment. The aMCI group performed significantly less well than controls on all but 6 of the neuropsychological measures that were administered. The mean ACE score of the aMCI participants who were exhibiting impairment on 3 or more episodic memory measures was significantly lower than for those showing impairment on 1 or 2 measures. Isolated memory impairment was rare.

Conclusions

Amnestic MCI is an important diagnosis in secondary and tertiary memory clinics. Memory impairments rarely present in isolation. There is scope to improve the efficacy and sensitivity of the clinical assessment of cognitive impairment in aMCI.

3.2.2 Introduction

Following publication of the initial set of criteria for MCI (Petersen et al. 1999), Petersen and colleagues (Grundman et al. 2004) proposed research criteria with a cut-off for verbal recall performance as objective evidence of episodic memory impairment. This approach has been challenged because it excludes patients who display exclusively visual episodic memory impairment (Alladi et al, 2006). Moreover, no justification for the choice of the outdated WMS-R version of the logical memory subtests is provided, and recent findings suggest that assessing episodic memory function in this manner may result in false positive MCI classifications (Brooks et al. 2007).

There is evidence to suggest that pure amnestic MCI (aMCI single domain) is rare; (Alladi et al. 2006;Hodges et al. 2006;Kramer et al. 2006;Tabert et al. 2006) and MCI case definition varies as a function of the neuropsychological tests used (Alladi et al, 2006; Loewenstein et al, 2006). It is not clear how MCI criteria might best be translated into clinical practice and there remains wide variability in terms of how aMCI criteria are applied, with regards the type and number of cognitive measures used to assess for impairment and the cut off points used to define impairment.

The second aim was to examine the diagnostic profile of aMCI patients referred to our tertiary assessment service and to evaluate a comprehensive battery of neuropsychological measures for its usefulness in this patient group. It was hypothesized that 1) a majority of MCI patients would demonstrate memory as well as other domains of cognitive impairment on our carefully selected neuropsychological assessment battery and that 2) classification of MCI participants (i.e. as normal, single or multi-domain amnestic or non-amnestic) would vary in accordance with the neuropsychological measures(s) and psychometric cut off points used to apply Petersen's MCI criteria and 3) a negative association between MCI patient's general level of cognitive functioning (as assessed by the total score obtained on the ACE and MMSE, and the number of impaired episodic memory measures was further predicted.

3.2.3 Method

A comprehensive neuropsychological battery of tests was administered to 46 patients with aMCI and 24 healthy elderly controls. Details of the participant groups and

neuropsychological assessment measures are provided in sections 2.1 & 2.2 of Chapter 2 Materials and Methods, respectively.

We present detailed neuropsychological baseline findings for 33 of the 45 patients who were identified in a retrospective analysis of referral types received between the months of September 2004 and April 2006, together with a further 16 MCI patients who were recruited through the Neuropsychological Assessment Service for Older People after this date (see Figure 3.1 Flowchart of referrals). All 46 aMCI subjects who consented to taking part in our study fulfilled inclusion / exclusion criteria for aMCI as outlined in Table 2.2. Twenty-four healthy elderly control participants were also recruited through a local dementia support group or were spouses or carers of patients who had attended the Neuropsychological Assessment Service, and met inclusion / exclusion criteria as outlined in Table 2.2.

3.2.3.1 Statistical Analysis

MCI patients were classified by subtype (i.e. single domain amnestic, multiple domain amnestic, non-amnestic single or multi) on the basis of their baseline performance on our study battery of neuropsychological measures, in accordance with both Criteria A (using a cut off of 1SD or more below control mean on two or more tests) and Criteria B (using a cut off of 1.5 SD or more below control mean on one or more test). We calculated z scores to determine where scores fell below the 10th percentile of control performances. Independent sample t-tests, with bonferroni corrections for multiple comparisions (and analysis of covariance where group differences in the demographic variables of age or FSIQ reached statistical significance), were carried out to compare group mean performances on each of the neuropsychological variables included in our study battery. As visual inspection together with a one-sample Kolmogorov-Smirnov goodness-of-fit test indicated that some of the data was not normally distributed, group comparisons were also conducted using the Mann-Whitney test.

In order to determine whether there is an association between general level of cognitive function and the consistency of episodic memory impairment, we divided the aMCI patients into two groups: those who displayed episodic memory impairment on 2 or fewer measures (where impairment was defined by a performance below the 10th%ile of the control group) and those who showed impairment on more than two measures. Their ACE and MMSE

scores were compared using an independent sample t-test. This analysis was repeated using the Mann-Whitney test (as visual inspection and one-sample Kolmogorov-Smirnov goodness-of-fit test indicated that the data were not normally distributed).

3.2.4 Results

From a group of 46 MCI patients recruited in this manner, 36 also performed at least 1 SD below our control mean on two or more of the episodic memory tasks that comprised the research battery, hence fulfilling Criteria A for aMCI. Thirty nine of the 46 participants met the more commonly applied psychometric cut off of a performance 1.5 SD or more below our age matched control mean on at least one episodic memory test from the research battery (fulfilling Criteria B for aMCI).

Four subjects exhibited no psychometric evidence of memory impairment relative to our control group on the study battery (despite having performed at least 1 SD below age norms on two or more episodic memory measures during their clinical neuropsychology consultation). Six MCI participants performed at least 1SD below the control mean on 1 episodic memory test from the research battery only.

When the non-episodic memory performance of the MCI participants (relative to the healthy elderly control group), was broken down in a similar manner, 32 participants were displaying impairments of 1SD or greater on at least 2 non-memory measures (in accordance with Criteria A), whilst 34 were exhibiting an impairment of 1.5 SD or more on 1 or more non-memory measure (in accordance with Criteria B).

Table 3.1 Breakdown of MCI subtypes following baseline neuropsychological testing in accordance with the use of 1 and 1.5sd cut off points

	Amnestic Single domain	Amnestic Multi-domain	Normal	Non-Amnestic Single or Multi-domain	Totals
Criteria A.	8	28	6	4	46
Criteria B.	8	31	4	3	46

Criteria A: performance of 1SD or more below age and IQ matched study control mean on two or more tasks; Criteria B: performance of 1.5 SD or more below and and IQ matched study control mean on one or more tasks

Table 3.2 Demographic data and performance of healthy controls (n=24) and amnestic MCI (n=46) on our neuropsychological test battery.

	Control (n=24) Mean (SD)	aMCI (n=46) Mean (SD)	Control v. MCI P
Age	70.8 (7.8)	73.9 (6.4)	0.080 ns
NART IQ	118.5 (3.3)	116.8 (7.6)	0.193 ns
MMSE (30)	29.0 (0.8)	28.4 (1.5)	P<0.01** ns adj
ACE			
ACE total (100)	94.5 (3.2)	89.5 (5.5)	P<0.001***
ACE Immediate recall (21)	20.3 (1.0)	20.2 (1.4)	0.291 ns
ACE delay (7)	6.4 (0.8)	4.3 (2.2)	P<0.001***
RCFT Scores			
REY copy (36)	34.2 (2.4)	34.0 (2.4)	0.378 ns
REY immediate recall (36)	18.9 (5.9)	13.0 (6.8)	P<0.01** ns adj
REY delayed recall (36)	17.5 (6.9)	11.6 (6.9)	P<0.01** ns adj
HVLT-R Scores			
HVLT-R total recall (12)	23.4 (5.1)	19.1 (4.8)	P<0.001***
HVLT-R delayed recall (12)	8.1 (2.7)	4.7 (3.3)	P<0.001***
HVLT-R DI (12)	9.9 (1.8)	8.3 (2.5)	P<0.01** ns adj
PAL 6 errors	7.9 (6.7)	17.0 (14.3)	P<0.01** ns adj
Trail Making Test			
TMT A (seconds)	40.2 (10.5)	45.2 (18.8)	0.238 ns
TMT B (seconds)	88.7 (30.7)	120.6 (68.7)	P<0.05* ns adj +0.06 ns
COWAT (FAS)	47.6 (14.3)	44.6 (11.7)	0.176 ns
Dual Task	93.3 (11.8)	94.2 (9.2)	0.361 ns
Boston Naming Test (60)	57.4 (2.9)	53.6 (5.3)	P<0.001***
Category fluency	51.4 (11.8)	36.4 (11.2)	P<0.001***
EENT (50)	49.0 (4.8)	45.1 (5.4)	P<0.01** ns adj
Graded Faces Test (30)	20.7 (3.6)	16.9 (5.0)	P<0.001***
Graded Naming Test (30)	23.8 (3.1)	20.7 (4.1)	P<0.01** ns adj

*** = p<.001, ** = p<.01, * = p<.05, (ns) = non-significant, (ns adj) = non-significant following Bonferroni correction for multiple comparisons

+ The p value for Mann-Whitney test is provided where a discrepancy in significance findings was observed across parametric and non-parametric methods of analysis and assumptions of equal variances or normal distribution were not met.

In table 3.1 it can be seen that in applying Criterion A (i.e. a performance of at least 1SD below age/IQ matched control mean on 2 or more measures), 8 could be classified as single domain amnestic MCI and 31 as multi-domain amnestic MCI. Of the remaining 7, 4 demonstrated no psychometric evidence of memory or any other domain of cognitive impairment relative to our healthy control group on the measures that formed part of the selected study battery, whilst 3 demonstrated impairments on two or more non-memory measures only.

Applying Criterion B (i.e. a performance of a least 1.5 SD below age/IQ matched control mean on 1 or more measures), 8 could be classified as single domain amnestic MCI and 28 as multi-domain amnestic MCI. Of the remaining 10, 6 demonstrated no psychometric evidence of memory or any other domain of cognitive impairment relative to our healthy control group on the measures that formed part of the selected study battery, whilst 4 demonstrated impairments on two or more non-memory measures only.

MCI and control participants did not differ significantly in terms of age $t(1.78) = 0.08$, $p>0.05$ or estimated pre-morbid IQ $t(-0.14) = 0.193$, $p >0.05$. These findings were replicated using non-parametric means of analysis (i.e. Mann-Whitney test). Despite a mean ACE score that exceeded suggested cut-off points for dementia (89/100) the MCI group performed at significantly lower levels on a number of neuropsychological measures that were administered including the ACE total and delayed story recall scores, the HVLT-R total and delayed recall scores, the BNT, GFT and the category fluency task. The findings were identical when the groups were compared using non-parametric methods of analysis (Mann-Whitney), with the exception of Part B of the Trail Making Test, where group differences in performance just failed to reach significance ($U=426$, $p=0.060$, $r=-0.186$) even prior to correction for multiple comparison.

All group means are based on numbers between 23-24 for controls and 44-46 for the MCI group, with the exception of the GNT, where data are based on 43 MCI and 19 control participants only. Baseline scores on the GNT were missing for 5 control and 3 MCI patients as this measure was added to the research battery shortly after study onset. Baseline RCFT data was also missing for one MCI participant, who declined to engage in this test as a

Table 3.3 Performance of patients with amnestic mild cognitive impairment on neuropsychological measures

Measure	Patients Performing below 10th percentile of control group performance n (%)
Episodic Memory	
HVLT_R total recall	11 (24)
HVLT_R delayed recall	22 (49)
HVLT_R discrimination index	17 (38)
Any HVLT-R measure	28 (62)
PAL 6 box errors	21 (46)
Rey Complex Figure Test delayed recall	14 (32)
Rey Complex Figure Test immediate recall	18 (41)
ACE immediate name and address recall	10 (22)
ACE delayed name and address recall	28 (61)
Patients with impairment on 1 measure only	5 (11)
Patients with impairments on 2 or more measures	35 (76)
Patients with impairments on 3 or more measures	26 (57)
Patients with impairments on 4 or more measures	17 (37)
Semantic Memory / Language	
BNT	17 (37)
EENT	13 (29)
GFT	18 (40)
GNT	15 (35)
Category Fluency	25 (57)
Participants showing impairment on one or more semantic memory measure	35 (76)
Attention / Executive Function	
TMT Part B	15 (33)
COWAT (FAS)	3 (7)
Dual Task	2 (5)
Visuospatial Function	
Rey Complex Figure Copy	4 (9)
Visuomotor Processing Speed	
TMT Part A	8 (17)

ACE, Addenbrookes Cognitive Examination; aMCI, amnestic mild cognitive impairment; BNT, Boston Naming Test; EENT, Edinburgh Exemplar Naming Test; GFT, Graded Faces Test; GNT, Graded Naming Test; HVLT-R, Hopkins Verbal Learning Test – Revised; PAL, CANTAB Paired Associate Learning Test; TMT, Trail Making Test; RCFT, Rey Complex Figure Test.

result of its challenging nature. The immediate and delayed components of the RCFT were not administered to one further MCI participant, for no apparent reason. One MCI participant refused to attempt the HVLT and as such 45/46 MCI baseline scores for the HVLT indices are available. Baseline scores on the EENT, GFT & Dual Task were not obtained for 1,1 & 2 MCI patients respectively, due to time constraints. Animal and single letter fluency data only (i.e. not the additional categories of fruits and vegetables or the letter F,A & S) were collected for a further 3 patients (2 and 1, respectively) for no apparent reason. NART estimates of pre-morbid IQ were not obtained for one each of our control and MCI participants and it is again unclear why the data was not collected in this case.

As in the Cambridge Memory Clinic study, not all aMCI participants demonstrated impairment across all episodic memory measures: 6 (13%) performed above the 10^{th}%ile of the control mean on all episodic memory measures, 5 (11%) showed impairment on a single test, 8 (17%) showed impairment on two memory measures and the remaining 27 (59%) were impaired on three or more tests.

Mean ACE scores for participants who were impaired on three or more episodic memory measures (M=87.54, SE=1.02) were significantly lower $t(44) = 0.003$, $p<0.01$ than for those showing impairment on two or fewer measures (M=91.95, SE=1.13). The difference in mean MMSE scores of the two MCI groups (M=28.04, SE=0.27; M=28.75, SE=0.35) was almost significant $t(44)=0.053$, $p>0.05$.

With reference to a performance falling below the 10^{th}%ile of our control group, just over 28 (61%) of our aMCI participants showed both verbal and visual episodic memory impairment. Although a significant proportion 11 (24%) demonstrated memory impairment of a verbal nature only, just 1 of our patients exhibited a pure visual memory deficit.

Only 6 (13%) of our aMCI patients exhibited an isolated impairment of episodic memory function. All but three of the other patients (i.e. 37/46) exhibited deficits in one or more additional domains of cognition, mostly that of semantic memory function, followed by attention and executive function.

3.2.5 Discussion

The applicability of Petersen's (Grundman et al. 2004) research criteria to clinical practice has recently been challenged on the grounds of exclusion of a significant number of patients who display episodic memory impairment of a visual nature only (Alladi et al, 2006). In contrast to the results of Alladi et al's (2006) study, around half of the aMCI subjects in our study showed impairment of verbal and visual memory, whilst all of the remaining subjects exhibited memory impairment of a verbal nature only. Only one of 46 MCI participants in our sample displayed an isolated impairment of visual episodic memory.

A recent study (Loewenstein et al, 2006) supports the notion that the type of episodic memory measure used may affect whether or not impairments are detected – a point we will return to later in the discussion. However, the absence of patients showing episodic memory impairment of a solely visual nature in our sample cannot be readily explained in terms of differential test sensitivities as near identical neuropsychological measures were used in previous studies (e.g. Alladi et al, 2006) to assess visual memory function. Only a small proportion of our aMCI patients demonstrated impairment on the visual episodic memory tasks per se. Administrative procedures might go some way to explain this observation. Specifically, our inclusion of an immediate Rey Complex Figure recall trial may have resulted in higher delay scores (Lezak et al, 2004), thus serving to reduce the sensitivity of this measure in our aMCI group.

Findings of studies examining patients who are 'at risk' of developing AD suggest that measures of verbal episodic memory are most sensitive to changes early in the disease course, followed by measures of visual memory (Collie and Maruff, 2000). It is therefore conceivable, taking into account our aMCI group's higher mean ACE score, that our aMCI sample contains a greater number of patients who are at an earlier stage of their disease course. Longitudinal follow-up, in particular observation of annual performances on these visual episodic memory measures, will determine whether this is indeed the case.

Recently, several studies have drawn attention to the substantial variability in MCI case definition as a function of the specific neuropsychological tests used (Alladi et al, 2006;Loewenstein et al, 2006). Consistent with this, in our study there was variability among aMCI patients as to which and how many episodic memory measures were impaired. This

finding was previously demonstrated (Alladi et al, 2006) and highlights the inherent difficulty in specifying the use of any single measure as a means of establishing impaired episodic memory function in aMCI. It would appear entirely reasonable and indeed a matter of good clinical practice to seek to establish consistency in performance across a range of episodic memory measures in defining aMCI and it will be of interest to see whether this is a significant determinant of outcome.

The variability in case definition of aMCI as a function of the cognitive measures employed, coupled with the inherent difficulties in specifying the use of a single common measure in the evaluation of aMCI, poses a major challenge for clinicians. Our findings would suggest that in employing Petersen's criteria for MCI, a patient could conceivably be classified as aMCI single domain, aMCI multiple domain, or 'worried well' depending on the cognitive measures that were employed. If the MCI subdivisions prove useful in a prognostic sense, the means by which the cognitive aspects of the criteria are put into operation by clinicians will require further clarification.

Mean scores on cognitive screening measures were significantly lower for subjects showing impairment on more than one episodic memory measure. This may reflect a more advanced disease course of this group. It is also possible that the single measure impairment group will prove to be a less stable one over time, with a number of patients returning normal neuropsychological profiles when re-tested. Alternatively, in cases where subjects show impairment on a single verbal memory measure only, this may have arisen secondarily to impairment in another cognitive domain, for example expressive language or attention/executive function (in which case the subject might be more accurately conceptualised as non-amnestic MCI). These possibilities and the prognostic implications of consistency and pervasiveness of impaired episodic memory performance(s) remain to be examined by way of longitudinal follow-up.

Our study adds to the growing body of evidence supporting the scarcity of a pure amnestic MCI syndrome (Kramer et al, 2006;Tabert et al, 2006;Alladi et al, 2006) and demonstrates that additional impairment often goes unnoticed unless participants undergo thorough neuropsychological assessment. Amongst 46 patients with aMCI only 6 patients (13%) presented with isolated memory impairment. This figure is well within the range of

previously reported rates. For example, Tabert and colleagues (2006) found that, following comprehensive neuropsychological assessment, only 19% of their aMCI cases were suffering from pure aMCI, whilst this figure reached 35% in Alladi et al's (2006) recent study.

It should be borne in mind, however, that the rate of cases with purely amnestic MCI will vary in accordance with how impairment is defined. For example, Kramer and colleagues (2006) showed that the number of cases that classified as pure aMCI was considerably higher (27%) when a cut-off of 1.5 standard deviation below the mean, as opposed to 1 standard deviation (resulting in a 5% rate), was used. It therefore remains a possibility that our less stringent definition of impairment (i.e. below the 10th percentile of healthy controls) may have resulted in an overestimation of the frequency of cases with non-pure aMCI. Identifying accompanying non-memory cognitive impairment nonetheless appears important in light of recent evidence indicating higher risk of conversion to AD in aMCI patients who show additional areas of cognitive impairment as compared to patients with pure aMCI (Tabert et al, 2006).

The results of the present study are also consistent with evidence indicating accompanying semantic memory impairment in aMCI (Alladi et al. 2006;Grundman et al. 2004;Hodges et al. 2006;Riberio et al. 2006;Tabert et al. 2006), with just 6 patients of 46 exhibiting episodic memory impairment in isolation, and 36 of the remaining 40 displaying evidence of semantic memory compromise. This finding may reflect an increased risk of conversion to AD from aMCI, although early semantic memory failure is by no means specific to AD (Clague et al. 2005;Graham et al. 2004), and whilst some studies report prognostic significance of performance on semantic memory measures (Artero et al. 2003;Blackwell et al. 2004;Estevez-Gonzalez et al. 2004), others have failed to do so (Beatty et al, 2002). The stage at which impairments in this domain become apparent does appear to vary in accordance with the sensitivity of the measure employed (Hodges et al, 2006). The intact performance of aMCI subjects on measures of lexical (letter) fluency, also previously reported (Alladi et al, 2006;Hodges et al, 2006), would suggest that the 'initiation' aspects of semantic fluency tasks do not pose any difficulty to aMCI patients.

In view of the sound mean performances of our aMCI subjects on cognitive screening measures (MMSE = 28/30; ACE = 89/100) it seems unlikely that consideration of such

scores will be of any value in ruling out the presence of additional domains of cognitive impairment. Reliance on clinical judgement to determine the presence/absence of additional domains of subtly impaired cognition, will similarly likely prove difficult when dealing with patients of above average pre-morbid IQ's who are performing at sound levels on cognitive screens. Taken together, the above observations raise the question of whether global screening measures coupled with clinical judgment are a sufficient means of investigating MCI, and if not, whether additional resources or an expanded skill base will be required to assess this population appropriately.

Our results reveal an absence of any significant difference in performance of aMCI and control groups on measures of visuo-spatial function and processing speed. In depth longitudinal evaluation of neuropsychological performance in MCI and Questionable Dementia (QD) suggests that visuo-spatial functions tend to fail secondarily to episodic memory and category fluency performances (Hodges et al, 2006;Perry and Hodges, 2000), although some heterogeneity is known to exist (Caine and Hodges, 2001). It is therefore once again possible that our failure to demonstrate group differences on a visuo-spatial copying task reflects an earlier disease stage of our aMCI sample. Alternatively, it is conceivable that varied and somewhat subjective scoring methodology for the Rey Complex figure copy task across different studies may be responsible for this finding.

Cross-sectional findings pertaining to visuomotor processing speed in MCI vary, with some studies reporting significant differences between MCI and control groups (Arnaiz et al. 2000;Grundman et al. 2004;Nordlund et al. 2005) and others, as ourselves, failing to do so (Albert et al. 2001;Crowell et al. 2002;Fox et al. 1998). The disparity in findings may simply reflect the heterogeneity of aMCI or alternatively disease stage. Group differences in processing speed might be more likely to exist where MCI samples contain significant numbers of patients in the pre-clinical stages of a subcortical dementia of a cerebrovascular nature. For example, there is some evidence to suggest a disproportionately strong association between perceptual speed and parkinsonian signs in MCI (Boyle et al, 2005).

It is noted that our aMCI sample was characterised by a high-average level of estimated pre-morbid general intellectual function, which of course introduces problems of generalising these findings other aMCI samples. Similar issues were present in a recent comparable study

(Alladi et al, 2006), although other socio-demographic characteristics (i.e. gender, ethnicity, education and occupation) were not reported, preventing further comparison between this and our study. There may therefore be a need to replicate these findings employing greater numbers of age and IQ matched healthy controls and aMCI patients with pre-morbid levels of intelligence together with other socio-demographic markers more closely resembling the general population mean.

3.2.6 Conclusion

In summary, MCI patients make up a significant number of referrals to ours and other older adult memory assessment services, with the most common MCI referral subtype in our sample being that of aMCI, followed by equal numbers of non-amnestic MCI and 'worried well'. Relatively few aMCI patients exhibit episodic memory compromise in isolation and fewer still show a visual but not verbal episodic memory deficit. Both the concept and criteria for MCI therefore appear to be a relevant and indeed a necessary adjunct to clinical practice.

The current findings highlight the inherent difficulties of specifying a single measure in the assessment of memory and other cognitive functions in MCI, whilst at the same time emphasising the need for clarification of the means by which MCI criteria can be put into operation clinically.

Initial attempts at better defining neuropsychological aspects of the aMCI criteria have been made (Grundman et al. 2004), however their application in a clinical sense remains inconsistent and their poor definition has not gone unnoticed (Portet et al, 2006). The existence of a number of neuropsychological measures of well documented sensitivity in aMCI and the strikingly similar mean performances of different clinic aMCI groups on such measures would suggest this need not be the case. Although the importance of exercising clinical judgment in arriving at a diagnosis of MCI cannot be ignored, it would nonetheless seem inevitable that further definition of the neuropsychological aspects of MCI criteria will be needed to facilitate identification of the MCI subtypes and to further our understanding of their respective prognoses.

4. Sensitivity and Specificity of Neuropsychological Measures to aMCI (Parts 4.1 – 4.5)

4.1 Cognitive screening in aMCI: A comparison of the ACE and MMSE.

(Published in part in the International Journal of Geriatric Psychiatry, 2009;

DOI: 10.1002/gps.2208)

4.1.1 Abstract

Background

Patients with mild cognitive impairment account for a significant number of referrals to old age psychiatry services and specialist memory clinics. The cognitive evaluation of such patients is commonly restricted to brief dementia screens, with no consideration to their suitability for assessing MCI.

Aims

To compare the abilities of two common screening instruments, the ACE and the MMSE, to detect MCI and early AD, and differentiate these conditions from cognitive changes that arise within the context of normal aging and depressive symptoms.

Method

The discriminative capacity of MMSE and ACE scores for 46 participants with MCI, 20 elderly out-patients with depressive symptoms, 20 early AD patients and 24 healthy age matched controls was determined by way of group mean comparisons and Regional Operating Characteristics (ROC) analyses. Healthy and depressive elderly control participants were examined seperately allowing us to test the robustness of previous findings purporting the ACE's specificity against depression. Where significant group differences existed on both screening measures, AUC, sensitivity, specificity, negative and positive predictive values were compared.

Results

The early AD participant group obtained significantly lower scores on both screening measures, in comparison with each of the other groups. MCI participants performed significantly less well than healthy age matched controls on the ACE, whereas the MMSE

failed to differentiate between these two groups. After adjusting for significant group differences in pre-morbid IQ levels, elderly patients with depressive symptoms could not be differentiated from the MCI or healthy control groups on the basis of their total scores on either screening test.

Conclusions

Whilst adequate to detect early AD using the higher cut off point (27/30), the MMSE lacks sensitivity to MCI. Comprehensive cognitive screening measures such as the ACE fare better, but still lack sensitivity to mild cognitive changes among the intellectually higher functioning elderly. In screening for early AD, but not MCI, both measures show adequate specificity against cognitive difficulties arising within the context of depressive symptoms.

The findings support the use of higher cut off values for the MMSE and ACE in screening for early AD whilst underscoring the need for fuller neuropsychological evaluation of cognitive function in high functioning elderly patients with MCI and depressive symptoms.

4.1.2 Introduction

Cognitive screening measures are widely used in clinical practice to help to determine the likely cause of a patient's cognitive complaints. To ensure evidence-based practice, it is necessary to know the diagnostic accuracy of the screening measure(s) being applied in this manner. Amongst the elderly, there are a number of conditions that may give rise to a loss of cognitive function. Of these, Alzheimer's disease (AD), (in its early and prodromal phases) and depression, are among the most common. Difficulties with differentially diagnosing each of these conditions are well established (Lezak 2004;O'Carroll et al. 1994).

Clinician surveys indicate that the Mini Mental State Examination (MMSE; (Folstein et al. 1975)) is the most commonly used cognitive screening instrument in clinical practice (Shulman et al. 2006). Two cut off values for the MMSE; 24/30 or 27/30 are usually applied. Their application varies in accordance with the educational background of the examinee and the relative importance of achieving high levels of sensitivity or specificity to dementia. Although the measure's sensitivity to AD is well established (Petersen et al. 1994), its ability to differentiate persons with aMCI from the healthy and depressed elderly is not.

Significant differences in the MMSE scores of these patient groups have been reported inconsistently and where they do exist, would appear to have little clinical meaning, ranging in magnitude from a minimum of less than one scale point (Ravaglia et al. 2005), to a maximum of just under 2 points (Slavin et al. 2007). As a majority of MCI patients score above the commonly used MMSE cut-offs (i.e. 24/30 and 27/30), there is considerable overlap in the scores of patients with MCI and age matched healthy controls. Recent findings suggest that a much higher cut off score that falls within what is typically viewed as a normal range i.e. 28/30, is required to achieve an adequate level of sensitivity to combined groups of highly educated dementia and MCI sufferers (O'Bryant et al. 2008).

Furthermore, there are a number of studies reporting low sensitivity values to MCI (Callahan et al. 2002;Ravaglia et al. 2005;Sager et al. 2006) as well as to the mild cognitive deficits that are known to accompany late life onset depression (Rajji et al. 2009). These findings

raise questions as to the appropriateness of screening with the MMSE where cognitive disturbances are mild.

In recognition of the important role of neuropsychological evaluation in facilitating early and differential dementia diagnoses, Mathuranath et al (Mathuranath et al. 2000) developed the Addenbookes Cognitive Examination (ACE). The ACE is a more comprehensive bedside cognitive screening measure, which incorporates all items of the MMSE together with items from several other well-established neuropsychological tests, with the effect of sampling a wide range of cognitive abilities. As with the MMSE, the ACE is relatively straightforward to administer requiring around 15 minutes.

The ACE has become an increasingly popular cognitive assessment tool in both clinical (Alladi et al. 2006;Bak et al. 2005;Dudas et al. 2005a) and research practice (Clague et al. 2005;Dudas et al. 2005b) within the UK. It appears to be sensitive to a relatively broad range of dementia presentations (Davies, 2007; Bak 2005, Mathuranath 2000) and the findings of several recent studies, among the author's own research group, suggest it may also be sensitive to cognitive deficits in MCI (Ahmed et al. 2008b;Alladi et al. 2006;Dudas et al. 2005a;Lonie et al. 2008).

In an initial validation study comprising 115 patients with dementia and 124 age and education matched controls, a cut off score of 83/100 on the ACE showed higher sensitivity, specificity and positive predictive power for dementia than the MMSE alone, at cut offs of 24/30 and 27/30 (Mathuranath et al. 2000). Another study has documented the test's specificity against major depression (Dudas et al. 2005a).

Sensitivity of the ACE to MCI and its specificity against affective disorders has not been replicated outside the author's (Mathuranath et al. 2000) research group. Furthermore, the mean age of the dementia and control groups was notably young in the initial validation study (i.e. 66.6 (8.9) for the dementia group, and 64.4(9.3) for the controls) as well as in the subsequent study examining specificity against affective disorders (AD group mean age = 68 years; NC = 64 years; Affective = 54.4 years). This raises the question of whether the findings from these studies are applicable to the 'older' old (i.e. 75 years and older), who

represent a far greater proportion of dementia sufferers and patients presenting to outpatient memory clinics.

In this study we therefore sought to compare the abilities of the MMSE and ACE to differentiate between groups of MCI, healthy and depressed elderly and early AD sufferers. We further sought to determine the discriminatory capacity of each screening measure at different cut off points, within a sample of elderly who more closely resemble the average age of persons attending geriatric outpatient clinics with cognitive complaints. On the basis of existing findings we predicted that the ACE, but not the MMSE, would be sensitive to MCI whilst retaining specificity against the cognitive effects of depressive symptoms within our comparatively 'older' participant sample.

4.1.3 Method

Groups of aMCI (n=46), Early AD (n=20), elderly patients with depressive symptoms (n=20) and healthy elderly control participants (n=24) completed measures of FSIQ (NART) and general cognitive ability (MMSE & ACE). The measures comprised part of a larger battery of neuropsychological tests administered at the initial baseline assessment of a 5 year longitudinal study on MCI. All participants therefore met inclusion and exclusion criterion for their respective groups, as specified in Chapter 2, Materials and Methods ,Table 2.2 Details of the method of subject recruitment together with psychometric and administrative characteristics of the neuropsychological measures employed can also be found within Chapter 2, Materials and Methods, sections 2.1 and 2.2.

4.1.3.1 Statistical Analyses

Group differences in NART FSIQ, age, MMSE and ACE were analysed using the Kruskal-Wallis Test as assumptions of normality and homogeneity of variance were not universally met (specifically the distribution of age scores for the early AD participants, FSIQ scores for the MCI group, and MMSE scores for the control, depression & MCI groups were abnormal). Furthermore, variances in FSIQ, ACE and MMSE scores across groups were non-homogenous. Where significant group differences were found, post hoc contrasts were carried out using the Mann-Whitney test with Bonferroni adjustment set at 0.05 / number of

contrasts performed i.e. ($p < 0.008$). Group differences in gender proportions were examined using the Chi-squared test.

The comparison of MMSE and ACE group means was repeated using a multivariate ANCOVA with NART FSIQ as a covariate, given the significantly lower mean FSIQ score of the AD group relative to both MCI ($p = 0.007$) and CT ($p = 0.002$) groups (see table 4.1.4, means (SD) of clinical and demographic data). The ANCOVA was followed up with pair wise comparisons using the Bonferroni method of correction for multiple comparisons. The ANCOVA and Kruskal-Wallis methods of analyses were inspected for consistency.

Where significant between group differences were present on both screening instruments, sensitivity and specificity values were generated using ROC analysis allowing for comparison of the discriminative utility of each screening test.

4.1.4 Results

Table 4.1 Means (SD) of clinical and demographic data

Variable	Controls (C;N=24)	Depressive symptoms (D;N =20)	aMCI (MCI;N=46)	Early AD (AD;N=20)	Statistic	Post hoc group differences (Mann-Whitney) p < 0.0083)
Age	70.8 (7.8)	74.0 (6.5)	73.9 (6.4)	75.3 (6.5)	H(3)=4.29,p=.32	
NART	118.5(7.8)	116.5 (6.7)	116.8 (7.6)	112.2(7.0)	H(3)=9.82,p=.02	C, MCI>AD
Gender	9 M: 15 F	5 M: 15 F	19M : 27 F	10 M:10 F	X=2.8;p=.43	
MMSE	29.0 (0.8)	28.5 (1.1)	28.4 (1.5)	24.3 (3.3)	H(3)=35.08,p=.00	MCI,C,D>AD
ACE	94.5 (3.2)	89.5 (6.6)	89.5 (5.5)	76.0 (9.0)	H(3)=46.46,p=.00	MCI,C,D>AD C>MCI C>D*

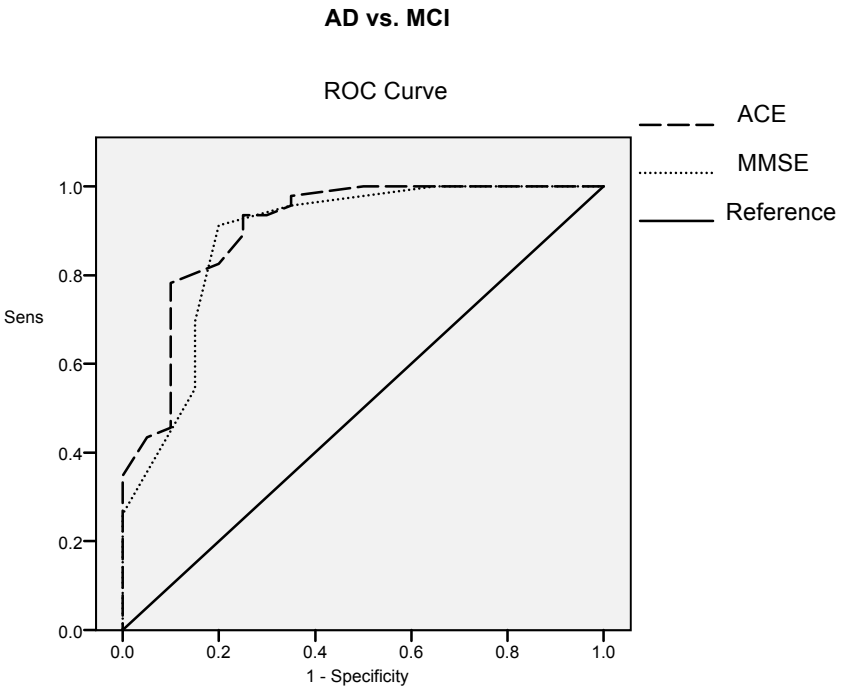
ACE, Addenbrookes Cognitive Examination; AD, Alzheimer's Disease, aMCI, amnestic Mild Cognitive Impairment; M, male; F, female; NART, national adult reading test; MMSE, mini mental state exam; * group differences were no longer significant following adjustment for pre-morbid IQ estimate (ANCOVA)

From table 4.1.4 it is apparent that the AD group performed significantly less well than all other groups on both screening measures (AD vs. MCI, ACE; U=85.0, r= -0.64, MMSE; U=106.5, r= -0.62; AD vs. C, ACE; U=5.5, r= -0.83, MMSE; U=34.5, r= -0.75; AD vs. D,

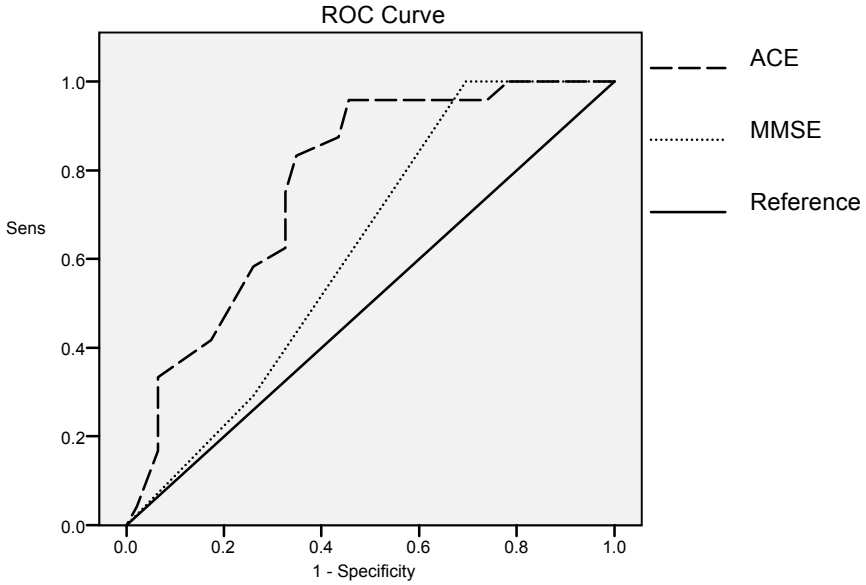
ACE; U=41.0, r= -0.68, MMSE; U=44.0, r= -0.68), where r represents the approximate effect size using the equation r = Z/square root N (Field 2005). Furthermore, the mean total ACE score for the MCI participant group was significantly lower than that of the normal elderly control group (U=253.0, r= -0.44), whereas mean total MMSE scores for these groups did not differ (U=416.5, r=-0.21).

The covariate NART FSIQ was significantly related to ACE and MMSE scores [F(1,103)=15.19, p <0.05, r=0.35; F(1,106=7.72, p<0.05, r=0.26]. The effect of participant group on total ACE and MMSE scores was also significant after controlling for NART FSIQ [F(3,106) = 29.39, p < 0.05; F(3,106) = 26.12, p < 0.05]. Pair wise comparisons revealed that group differences in mean total ACE scores for the Control and Depression participants (i.e. Depression < CT) no longer reached significance once the effects of FSIQ were controlled for t(41) = 2.46, p = >0. 5. Findings across the two methods of analysis did not otherwise differ.

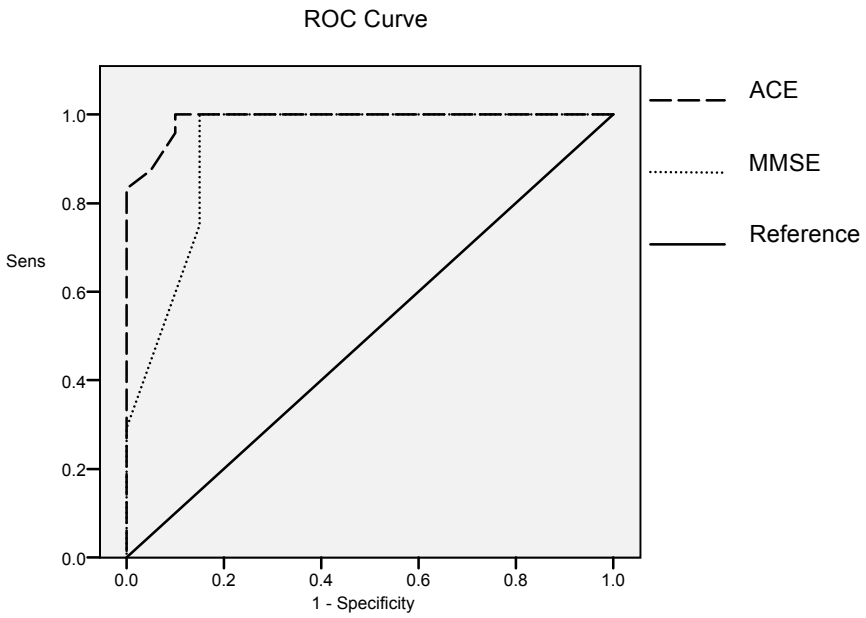
Figure 4.1 ROC curves comparing the abilities of the ACE and MMSE to discriminate MCI and AD participants from controls and each other.



MCI vs. CT



AD vs. CT



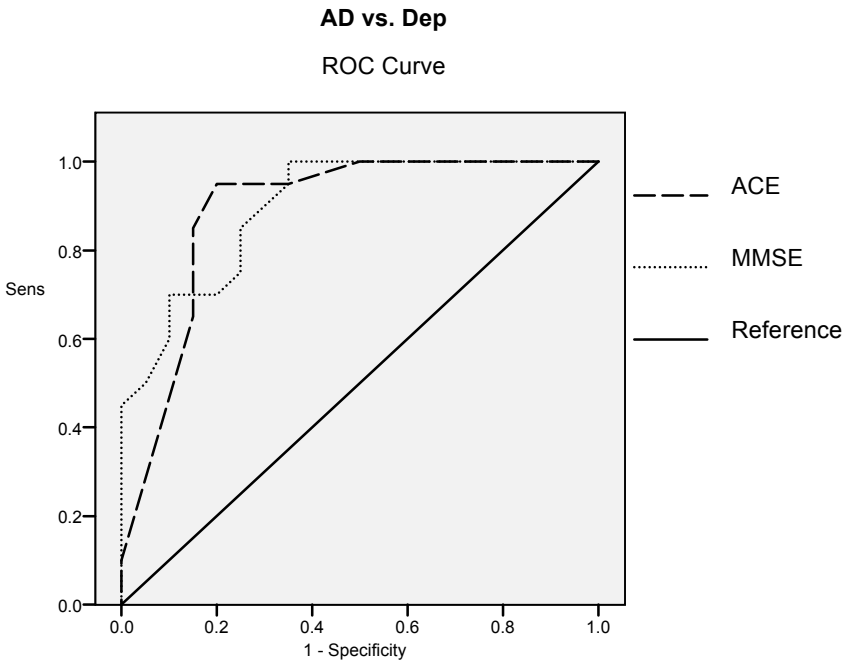


Figure 4.1 shows mean total scores on both the ACE and MMSE discriminate the early AD patients from all other participant groups with similar levels of accuracy and demonstrates that the ACE has greater combined sensitivity and specificity in distinguishing healthy elderly from MCI participants than the MMSE, for which discriminative capacity fails to reach significance.

Table 4.2 Positive and Negative Predictive Values for differing ACE and MMSE cut off points as a function of participant group

Screening Measure / cut off point Predictive value	Early AD vs. Combined control groups PPP	NPV	MCI vs. Combined control groups PPP	NPV
MMSE 24/30	100	77.78	100	49.44
MMSE 27/30	85	57.14	82.35	56.16
ACE 83/100	78.26	88.65	50	48.75
ACE 88/100	66.67	94.59	70.97	59.32

ACE, Addenbrookes Cognitive Examination; AD, Alzheimer's Disease; MCI, Mild Cognitive Impairment; MMSE, Mini Mental State Examination; NPV, Negative Predictive Value; PPV, Positive Predictive Value

Table 4.3 Validity of cognitive screening measures in differentiating AD and MCI from the healthy and depressive elderly at two commonly applied cut off points

Group Comparison	Screening measure	Cut-off	Sensitivity %	Specificity %	AUC	Sig
AD vs. MCI	MMSE	24	50	98	0.88	*
		27	80	91		
	ACE	83	75	89	0.91	*
		88	90	55		
AD vs. D	MMSE	24	50	100	0.89	*
		27	80	95		
	ACE	83	75	75	0.90	*
		88	90	65		
AD vs. C	MMSE	24	50	100	0.93	*
		27	80	100		
	ACE	83	75	100	0.99	*
		88	90	96		
MCI vs. C	MMSE	24	2	100	0.62	n.s.
		27	9	100		
	ACE	83	11	100	0.77	*
		88	48	96		

ACE, Addenbrookes Cognitive Examination; AD, Alzheimer's Disease; AUC, Area Under Curve; D, Depression; C, Control; MCI, amnestic Mild Cognitive Impairment; MMSE, mini mental state exam; n.s., non-significant $p>0.05$; *, sig $p<0.05$; PPV, Positive Predictive Value; NPV, Negative Predictive Value.

Figure 4.2 Total score on the ACE as a function of participant group

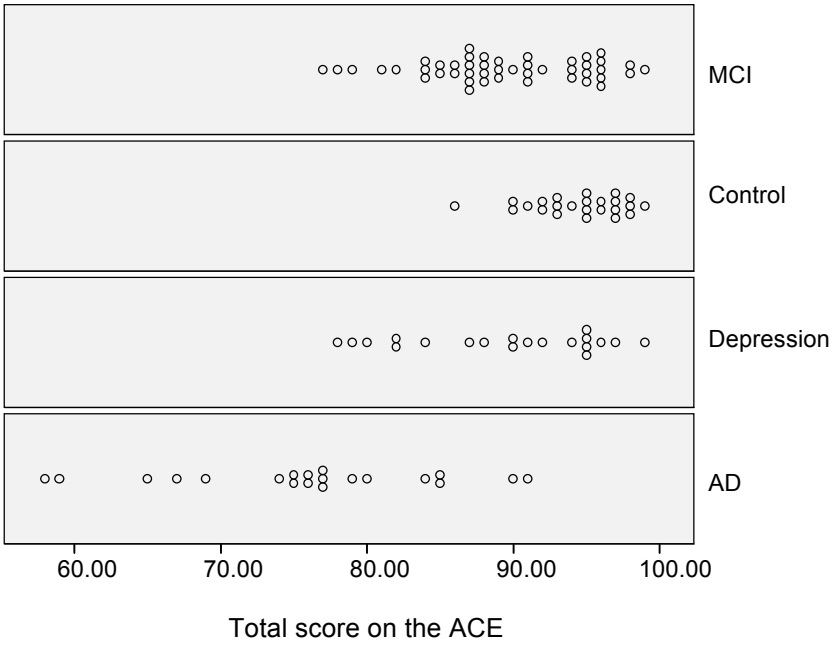


Figure 4.3 Total score on the MMSE as a function of participant group

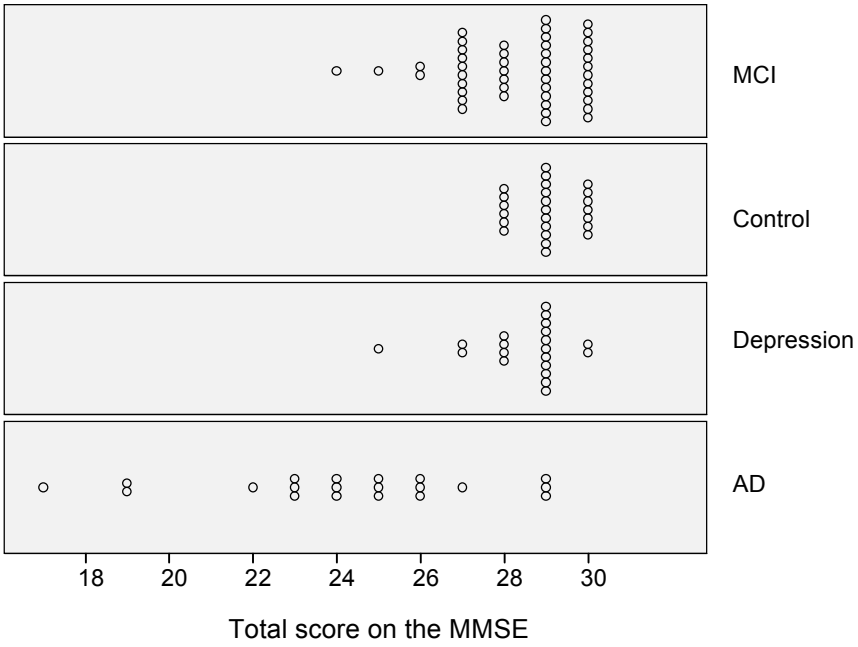


Table 4.3 displays the sensitivity, specificity and areas under the curve (AUCs) relating to the two most commonly applied cut off values for each screening measures. An increase in sensitivity and loss of specificity is universally observed in conjunction with the use of the

higher cut off values (with the exception of the MMSE where cut off scores of both 24/30 and 27/30 retain 100% specificity in differentiating AD and MCI from healthy controls in our sample). Satisfactory (i.e. >80%) sensitivity and specificity combinations are observed for the differentiation of 1) early AD patients from elderly with depressive symptoms using the higher MMSE cut off of 27/30; 2) early AD patients from healthy controls using the higher cut off point for the MMSE 27/30 or the ACE 88/100; and 3) early AD patients and MCI participants using the higher cut off point for the MMSE 27/30. The sensitivity of both the MMSE and the ACE (at high and low cut off points) to MCI is low, at the expense of high specificity against normal aging. The highest overall discriminative value (AUC = 0.99) relates to the ability of the ACE to differentiate between early AD patients and the healthy elderly. Good levels of overall discrimination in differentiating AD from elderly patients with depression and those with MCI (AUC = 0.89; 0.88 respectively) are also seen in association with this screening measure.

As expected, the highest negative (NPV) and positive predictive values (PPV) were observed for the classification of AD patients amongst combined depressive and normal elderly controls, where PPV represents the percentage of participants scoring below a designated level with a diagnosis of AD and NPV represents the percentage of participants scoring above this level without an AD diagnosis. More specifically, the highest PPV was observed in association with the lower boundary MMSE cut off score of 24/30, where 100% of persons performing below this level had a diagnosis of early AD. NPV's were highest for the higher cut off value of the ACE, where 94% of persons scoring above this level were either normal elderly or depressive elderly control participants as opposed to early AD patients.

In relation to the classification of MCI participants amongst normal and depressive elderly controls, PPV's were highest for the MMSE where 82% of persons obtaining a score of less than 27/30 and 100% of those scoring less than 24/30 met criteria for MCI. NPV's were universally low (i.e. <60%).

Figures 4.2-4.3 illustrate the limited overlap in total ACE and MMSE scores obtained amongst the early AD and healthy elderly control participants. By contrast, a considerable

degree of overlap in scores on both cognitive screening measures is apparent for the remaining groups.

4.1.5 Discussion

Our findings indicate that both the MMSE and the ACE differentiate early AD patients from elderly patients with depressive symptoms, healthy elderly controls and patients with MCI, with a similar degree of overall accuracy. Similar levels of sensitivity and specificity to dementia have been reported previously for both the MMSE (at a cut off of < or = 24/30, sensitivities = 0.52, 0.49, 0.52 and specificities =100, 0.96, at a cut off of < or = 27/30, sensitivities = 0.8, 0.83, 0.74 and specificities =100, 0.72, 0.96) and the ACE (at a cut off of < or = 83/100, sensitivities =0.75, 0.87, 0.82 and specificities =100, 0.71, 0.96) at a cut off of < or = 88/100, sensitivities = 0.90, 0.98, 0.93, 0.93 and specificities = 0.96, 0.59, 0.71, 0.82 (Bier et al. 2004;Bier et al. 2005;Dudas et al. 2005a;Mathuranath et al. 2000).

One exception is noted. Specificity levels are greater in ours and the study by Mathuranath et al., (Mathuranath et al. 2000) in comparison to Bier et al., (Bier et al. 2004;Bier et al. 2005). This is likely to reflect the greater heterogeneity and overall pathological loading of Bier et al's control group (which is reported to include patients with MCI, vascular lesions, progressive isolated memory impairment, mood disorders as well as cognitive impairments as a result of toxic exposure). It is not clear whether this control group was matched to that of the dementia group for age, education or pre-morbid IQ, which further complicates cross study comparison of findings. Finally the mean age of the control group is somewhat lower than that of our own (i.e. 67.3 years compared to 72.4 years) in this study.

The MMSE was associated with higher positive predictive values for dementia and MCI than the ACE. This finding contrasts that of Mathuranath at el (Mathuranath et al. 2000) who report 'similar specificities to dementia but poorer sensitivity and predictive values for the MMSE. Close inspection of the table of PPVs, however, suggests that at higher i.e. 40% base rates, (which represent levels likely encountered within specialist memory clinic settings such as our own), there is little difference in the PPV of these instruments with the exception of a lower PPV in association with the higher cut off point for the ACE. The latter observations are in keeping with our own findings, and the lower PPV values reported overall in our own study likely reflect the higher general level of cognitive functioning of our

dementia group (i.e. ACE 76/100 as compared to 64/100), and/or the inclusion of a wider range of dementia diagnoses in the Mathuranath study.

Our findings together with previous results (Mathuranath et al. 2000) suggest that clinicians can be confident that where MMSE scores fall below 24/30 among those with above average pre-morbid intellect, cognitive impairment is likely to represent an Alzheimer or other dementia process as opposed to cognitive changes that relate to mild-moderate depressive symptoms or the normal aging process (PPV at a base rate of 40% range from 0.89 – 100%). Conversely, in accordance with the present findings, a score of above 88/100 on the more comprehensive ACE, equally likely signifies the absence of a clinically diagnosable AD process (PPV = 0.95).

Although the ACE shows greater sensitivity to MCI than the MMSE, overall sensitivity levels were low in the present study. Previous research examining the discriminatory abilities of the more recent version of this test (ACE-R) also reports significant differences in mean group performances of the MCI participants (n=23) relative to the Control and AD groups (n=23 for both). However, sensitivity and specificity values were not reported in association with these findings (Mioshi et al. 2006) and the authors acknowledge 'it is not clear whether these findings would apply equally to an older patient group'.

Together with the low NPV for the MMSE and ACE at both higher and lower cut off values, the low sensitivity values reported in the present study suggest that a large proportion of elderly patients who report cognitive complaints yet score above ACE and MMSE cut off points will demonstrate evidence of cognitive compromise relative to their age/IQ contemporaries on formal measures of their neuropsychological functioning. Indeed, our findings would suggest that around half of all elderly patients who fulfil Petersen's criteria for MCI (Petersen et al. 1999) may fall into this category. These patients are therefore likely to be diagnosed as worried well or cognitively normal where screening instruments are used as the sole means of investigating their cognitive complaints.

Unlike Dudas et al (Dudas et al. 2005a) who reported a significantly lower mean total ACE score for a group of 60 elderly patients with affective disorders relative to 127 elderly controls, no differences in the total mean ACE scores of elderly patients with depressive

symptoms and healthy controls were found following adjustment for FSIQ in the present study. Whilst the mean scores on the ACE were similar across the two studies (i.e. Affective / depression groups = 89/100 / 89.5/100; Control groups = 93.9/100 / 94.5/100) the mean ages of both participant groups in Dudas et al., (Dudas et al. 2005a) study were notably younger than in our study (i.e. Affective/depression groups = 54.4 years/ 74 years; Control groups = 64.4 years / 71 years) and it is not clear that the significant differences in the ages of the affective and control groups were accounted for. Furthermore, the authors comment that the absolute difference in the mean ACE scores for the affective and control groups was very small (i.e. 89 + -8.2 vs. 93.9 + -3.5, respectively). It is therefore possible that the inclusion of NART FSIQ as a co-variate (in our own analysis), together with the smaller number of participants used, resulted in a loss of power. As close to half of the patients included in the affective group of the Dudas et al., (Clague et al. 2005) study, as well as our own, met criteria for major depression, our negative findings do not appear to be accountable for, in terms of variability, in levels of depressive symptoms across studies.

4.1.6 Conclusion

Overall scores on both the ACE and MMSE discriminate patients with AD from normal elderly controls, MCI patients and elderly patients with depressive symptoms with little difference in rates of overall accuracy. The MMSE is insensitive to Mild Cognitive Impairment amongst the elderly with high average pre-morbid ability. Even with the use of more comprehensive bedside cognitive screening instruments such as the ACE, our ability to detect cognitive impairments that do not amount to a frank dementia remains unacceptably low, implying that large numbers of elderly patients presenting with memory complaints who are cognitively evaluated in this manner may be mistakenly categorised as ‘worried well’ or ‘normal for age’. The ACE and MMSE appear to be relatively specific against the known cognitive effects of depressive symptoms, although this finding would benefit from replication in a larger group of patients with moderate to severe levels of depressive symptoms matched for age and pre-morbid IQ.

4.2 A comparison of free and cued episodic memory paradigms in the differential diagnosis of MCI and early AD

4.2.1 Abstract

Background

Neuropsychological assessment, in particular evaluation of episodic memory, is one of the most sensitive means of detecting cognitive impairment and identifying individuals in the early and pre-clinical stages of Alzheimer's disease. Despite this, there is little information to guide the selection of memory paradigms in the assessment of these conditions with tasks usually chosen independently of knowledge of the neuropathogenisis and cognitive neuropsychology of AD.

Aims

To compare the abilities of cued and free recall episodic memory measures to discriminate between MCI and early AD patients and the healthy or depressive elderly. It was hypothesized that 1) a computerised cued recall measure of cross modal associative learning would better discriminate between the above patient groups than more traditional paper and pencil memory tasks 2) free recall tasks would be more sensitive to memory failure in AD and MCI than cued recall paradigms and 3) cued recall measures would show greater specificity against depression and normal aging than measures of free recall ability.

Method

The discriminative capacities of 2 cued (1 verbal and 1 visual) and 4 free recall measures (3 verbal and 1 visual) for groups of 46 participants with MCI and 20 with early AD relative to a combined control group comprising 20 elderly out patients with depressive symptoms, and 24 healthy age matched elderly, were determined by way of group mean comparisons. As the focus here was on the practical application and selection of memory measures in the clinical detection and differential diagnosis of AD and aMCI, as opposed to exploring the specific cognitive effects of *depressive symptoms* on memory function, the healthy and depressive elderly control groups were combined for the present analysis. Where significant group differences were found, AUC, sensitivity and specificity values were computed by way of

ROC analyses and percentages of overlapping scores for each of the 5 memory measures were calculated.

Results

The early AD and MCI participants performed more poorly than the combined control group on all episodic memory measures other than PAL, where the difference in error rates between the control and MCI groups just failed to reach significance (p=0.051).

The discriminative capacities of each of the six memory measures in differentiating early AD patients from the combined control groups were consistently high, ranging from a minimum of AUC 0.904 for HVLT-R delay to a maximum of 0.943 for PAL. AUC's were moderate for the MCI vs. control comparisons (AUC 0.6 PAL – 0.7 ACE and HVLT-R delay).

Each of the 6 memory measures was highly sensitive to AD and highly specific against normal aging and depressive symptoms. None of the 6 memory measures was associated with high levels of sensitivity and specificity in discriminating MCI from the combined control groups.

When either specificity or sensitivity was pre-set at the level of 0.80, free recall measures provided a marginally better balance of these two indices than cued recall measures, in differentiating both MCI and AD from the combined control groups.

Conclusions

Both cued and free recall memory paradigms differentiate early AD from normal aging and cognitive difficulties associated with depressive symptoms with high levels of sensitivity and specificity. At a designated minimum 'acceptable' level of sensitivity (i.e. 0.8), cued recall measures were not associated with higher levels of specificity in discriminating early AD or MCI participants from controls. None of the memory measures were adequately sensitive and specific to MCI, although free recall measures showed the best combination of these properties. These findings support the use of free recall as opposed to cued recall measures

in the early differential diagnoses of AD and MCI whilst highlighting the limitations of existing memory measures in differentially diagnosing MCI.

4.2.2 Introduction

Episodic memory impairment is the hallmark of aMCI, pre-clinical and early AD. The importance of diagnosing AD accurately, at an early stage, is widely recognised and among other advantages facilitates patient access to treatment and appropriate forms of support. Neuropsychological assessment, in particular evaluation of episodic memory, is one of the most sensitive means of detecting cognitive impairment and identifying individuals in the early and pre-clinical stages of Alzheimer's disease (Lowndes and Savage 2007). Consequently, memory assessment in the early and differential diagnosis of AD has formed the focus of a large number of research studies (Ahmed et al. 2008b;Clague et al. 2005;Fowler et al. 1995;Fowler et al. 1997;Fowler et al. 2002;Graham et al. 2004;Greene et al. 1996;Lee et al. 2003;Scahill et al. 2005;Swainson et al. 2001).

Episodic memory tasks vary on a number of levels such as, the nature of the material to be learned i.e. verbal, visual or spatial (or any combination of these), the type of learning required i.e. associative (i.e. the pairing or linking of information) or non-associative (i.e. single word list learning), and the extent to which recall is supported or cued i.e. recognition or free recall. There is evidence to suggest that variability in these aspects of memory tasks can influence the diagnostic performances of episodic memory measures (Lowndes and Savage 2007;Pike and Savage 2008).

The sensitivity of delayed free recall paradigms to aMCI (Alladi et al. 2006;Lonie et al. 2008), pre-clinical (Collie and Maruff 2000;Grober and Kawas 1997) and early stage AD (Bondi et al. 2008) is well established, and a majority of studies investigating the longitudinal course of aMCI have employed one or more delayed free recall paradigm as the primary measure of episodic memory function. Furthermore, among patients fulfilling Petersen's criteria for aMCI (Petersen et al. 1999), effect sizes, where these are reported, are generally largest for measures of free relative to cued recall (or other non-memory tasks) (Alladi et al. 2006;Backman et al. 2005;Lonie et al. 2008). Other studies have likewise reported greater sensitivity of free relative to cued recall memory paradigms in association with early AD (O'Connell et al. 2004) and pre-clinical AD sufferers (Fowler et al. 1995).

Whilst sensitive to early memory failures of a pathological basis, free recall measures lack specificity against depression (Lichtenberg et al. 1995;Zakzanis et al. 1998), and other non-AD forms of dementia (De Jager et al. 2003).

Dierckx et al (Dierckx and Engelborghs 2007) therefore proposed the use of a delayed cued recall measure in differentiating between early AD, MCI and elderly controls with and without depression. The underlying assumption was that by minimising reliance on the executive aspects of memory performance, greater specificity to AD, and MCI, as it represents pre-clinical AD, could be achieved. Using this approach and applying a cut of value of 8 / 12 for the combined total number of words and object location pairings recalled following a delay and with cueing, high levels of sensitivity to AD (83%) and specificity against depression (85%) were achieved. Sensitivity to MCI was, however, considerably lower (53%), and the authors attributed this to the heterogeneous nature of the MCI participant group, only a portion of whom would likely progress to AD over time.

Whilst several additional studies have reported favourable performances of cued recall measures in discriminating between groups of depressive, MCI, NC and early AD patients (Bennett et al. 2006;Ritter et al. 2006;Swainson et al. 2001) others have reported insufficient sensitivity of cued recall measures to MCI and pre-clinical AD, relative to healthy age controls (Arnaiz et al. 2000;Crowell et al. 2002;Dudas et al. 2005b;Godbolt et al. 2005;Hudon et al. 2006;Westerberg et al. 2006).

Attention has also been drawn to the importance of incorporating knowledge of the neuropathogenesis of AD into the selection of memory paradigms for its early stage assessment (Fowler et al. 1995;Lowndes and Savage 2007;Pike and Savage 2008). It has been proposed that the cross modal (i.e. pairing of visual and spatial information) associative learning requirements of the PAL from the CANTAB battery closely reflect the role of medial temporal lobe structures, where new information is bound together in order to encode the complex relational structure of personal experiences (Fowler et al. 1997;Swainson et al. 2001). For this reason, the PAL subtest from the CANTAB battery, which requires pairing of visual and spatial information, is thought to represent a valuable paradigm for the assessment of episodic memory function in early AD.

The sensitivity of PAL to memory failure in early AD is well established (Fowler et al. 1995;Lee et al. 2003;Swainson et al. 2001), although there is little information regarding it's differential diagnostic performance in relation to other, more traditional paper and pencil delayed free recall measures (O'Connell et al. 2004). 100% specificity values were reported for a small sample of 34 AD and 16 healthy age matched controls (O'Connell et al. 2004). High specificity values have also been reported in relation to a sample of depressed elderly (Swainson et al. 2001) and patients with diagnoses of frontal and semantic variant FTLDs (Lee et al. 2003). Several studies have also reported higher numbers of errors made across the learning trails of PAL in association with MCI (Ahmed et al. 2008b;Alladi et al. 2006;Collie et al. 2001;Egerhazi et al. 2007;Lonie et al. 2008;Petersen et al. 1999) although the predictive validity of this heightened error count remains controversial (Ahmed et al. 2008b;Blackwell et al. 2004;Fowler et al. 2002). The specificity of PAL against depression has not been replicated outside the author's research group.

Memory assessment is a key component in the diagnosis of early AD and MCI, where performance on bedside cognitive assessment can remain inside normal limits. There is considerable scope for variation in the application of memory tasks within this context, which has the potential to affect the diagnostic contribution of cognitive assessment to the early and differential diagnosis of AD and MCI. In clinical practice, the key challenge lies in differentiating memory complaints arising in the context of a depressive illness or as part of the normal aging process, from those that represent a prelude to dementia. By contrast, relatively little importance is attached to detailed investigation of memory impairment accompanying symptoms of depression in the elderly. As such, comparative performances of combined groups of healthy and depressive elderly, with MCI and early AD patients are of primary importance in a clinical sense.

In this study we therefore sought to determine 1) whether sufficient levels of sensitivity to MCI and AD could be retained alongside improved specificity against normal aging and depression with the use of cued as opposed to free recall paradigms to assess episodic memory function. We further sought to 2) compare the diagnostic performance of the PAL to that of the more traditional non-computerised and non-associative episodic memory tasks in differentiating early AD and MCI from combined healthy and depressive elderly.

4.2.3 Methods

Groups of healthy elderly (n=24), depressive elderly (n=20), early AD (n=20) & MCI participants (n=46) were administered several measures of episodic memory function including The Hopkins Verbal Learning Test - Revised, The Rey Complex Figure Test, The Paired Associated learning subtest from the CANTAB and the delayed recall component of the name and address task from the Addenbrookes Cognitive Examination (ACE) together with a measure of pre-morbid IQ (NART) and two cognitive screening measures (MMSE & ACE). These measures were administered as part of a wider neuropsychological test battery.

Details of the method of subject recruitment, participant characteristics together with psychometric and administrative characteristics of the cognitive screening measures employed can also be found within Chapter 2, Materials and Methods, sections 2.1 and 2.

4.2.4 Statistical Analysis

Differences in gender proportions between participant groups were examined using the chi-square test. Differences in the mean age of the participant groups were analysed using One-way Anova. Visual inspection of the data together with Kolmogorow-Smirnov tests revealed that the distributions of NART estimated FSIQ scores for the combined (healthy and depressive) control groups and MCI participants were not normal. Group differences in NART FSIQ scores were therefore examined using non-parametric methods i.e. the Kruskal-Wallis test followed by Mann Whitney post hoc comparisons comparing NART estimated FSIQ scores of the MCI and early AD groups with that of the combined depressive and healthy elderly control groups with Bonferroni level of correction set at $p<0.025$ accounting for the two comparisons.

As the NART estimated FSIQ score was significantly lower for the early AD group relative to the combined controls, an ANCOVA was carried out with NART FSIQ as covariate to examine group differences in scores on each of the 6 memory measures. Group comparisons in performance on the 6 memory measures were repeated using the Kruskal-Wallis test followed by Mann-Whitney post hoc comparisons (with the Bonferroni level of correction set at $p<0.004$) as assumptions of normality and homogeneity of variance were not universally met in either of the two participant group comparisons (i.e. CT vs. MCI; CT vs. AD).

Although the NART estimated mean FSIQ scores of the MCI and combined control groups were not significantly different, assumptions of normality and homogeneity of variance were not universally met. The Kruskal-Wallis test was therefore carried out to test for group differences in scores on each of the 6 memory measures. This was followed up with group comparisons using the Mann-Whitney procedure (with the Bonferroni level of correction set at $p<0.004$) as described above.

ROC analyses were carried out for both group comparisons (i.e. AD vs. Dep + CT & MCI vs. Dep + CT) and all memory measures. Sensitivity values were set at 0.8 and corresponding specificity and cut off values were reported for each of the memory measures and both group comparisons. With specificity values set at 0.8, corresponding sensitivity and cut off values were reported for the two group comparisons on all memory measures.

Table 4.4 Episodic memory and demographic data for Early AD, MCI and combined healthy and depressive controls

	Early AD (n=16-20)		MCI(n=44-46)		Healthy and Depressive Controls (n=43-44)	
	Mean(SD) Sig of F-value	Median(range) Sig of U-value	Mean(SD) Sig of F-value	Median(range) Sig of U-value	Mean(SD)	Median(range)
NART FSIQ	112.20 (1.57)	110 (99-124) **	116.78 (1.13)	120.00(92-126)	117.56 (0.08)	118 (100-126)
Age (years)	75.30 (1.46)	78.00 (64-86)	73.89 (0.95)	74.50 (58-85)	72.25 (1.12)	72.00 (59-85)
Gender (M/F)	10:10		27:19		30:14	
PAL 6 box errors	39.69 (2.79)***	42.50(13-50)***	17.04 (2.10)	14.50 (0-49)	11.77(1.61)	9.00 (0-50)
HVLT DI	4.63 (0.68)***	5.00 (-2-11)***	8.33 (0.38)**	8.00 (1-12)***	9.61(0.31)	10.00 (3-12)
HVLT Total	12.79 (1.16)***	14.00 (4-22)***	19.11(0.72)**	19.00 (10-31)**	22.05 (0.90)	21.50 (9-34)
HVLT Delay	1.42 (0.64)***	0.00 (0-10)***	4.71(0.50)***	5.00 (0-11)***	7.39 (0.46)	8.00 (0-12)
RCFT-Delay	2.65 (0.85)***	0.00(0-14)***	11.55 (1.03)**	11.50 (0-28)**	15.84 (1.05)	16.00 (0-31)
ACE-Delay	1.35 (0.45)***	0.00(0-7)***	4.30 (0.32)**	5.00 (0-7)***	5.75 (0.26)	6.00 (0-7)

Significance of difference in mean score relative to combined control mean - * $p<0.05$; ** $p<0.01$; *** $p<0.001$; FSIQ = Full Scale Intelligence Quotient; M=male; F=female, PAL=Paired Associate Learning Test; HVLT = Hopkins Verbal Learning Test - Revised; DI=Discrimination Index; Total=total number of words recalled across the three learning trials; Delay=total number of words recalled following a delay; RCFT=Rey Complex Figure Test; ACE=Addenbrookes Cognitive Examination

4.2.5 Results

Table 4.5 Sensitivity and specificity values with corresponding cut off values for patient and control group comparisons on the episodic memory measures

Memory Measure Group Comparison (AUC)	Sens > or = 0.8 (cut off score)	Spec	Spec > or = 0.8 (cut off score)	Sens
Cued Recall Measures				
PAL				
AD vs. CTs (0.94)	0.83 (18.5)	0.875	0.812 (34.5)	0.93
MCI vs. CTs (0.60)	0.837 (17.5)	0.391	0.848 (3.5)	0.209
HVLT-R DI				
AD vs. CTs (0.92)	0.841 (7.5)	0.842	0.842 (7.5)	0.841
MCI vs. CTs (0.64)	0.841 (7.5)	0.349	0.907 (11.5)	0.182
Free Recall Measures				
HVLT-R delay				
AD vs. CTs (0.90)	0.818 (4.5)	0.895	0.842 (3.5)	0.886
MCI vs. CTs (0.71)	0.818 (4.5)	0.465	0.837 (8.5)	0.386
RCFT delay				
AD vs. CTs (0.94)	0.818 (10.5)	0.947	0.842 (6.5)	0.886
MCI vs. CTs (0.67)	0.818 (11.2)	0.512	0.814 (18.2)	0.341
ACE delay				
AD vs. CTs (0.92)	0.864 (4.5)	0.947	0.842 (3.5)	0.886
MCI vs. CTs(0.72)	0.864 (4.5)	0.488	0.837 (6.5)	0.455
HVLT-R total				
AD vs. CTs (0.88)	0.818 (18.5)	0.789	0.895 (16.5)	0.727
MCI vs. CTs(0.66)	0.818(16.5)	0.244	0.8 (22.5)	0.477

Spec = the highest value associated with a sensitivity value of 0.8 or above; Sens = the highest value associated with a specificity value of 0.8 or above

There was no significant association between participant group and gender, chi-squared (2) = 2.07, $p>0.05$. The mean ages of the participant groups did not differ, $F(2, 107)=1.51$, $P>0.05$. NART estimated FSIQ scores for the MCI and combined control groups did not differ ($U= 924.5$, ns, two tailed, $r=-0.04$) whereas the NART estimated FSIQ score for the AD group was significantly lower than that of the combined controls ($U= 231.00$, $p<0.01$, two tailed, $r=-0.37$).

ANOVA findings revealed that the mean performance of the combined control group was significantly higher than that of MCI participants on all memory measures (HVLT-R DI, $F(1, 87) = 6.92$, $p=0.01$, $r=0.27$; HVLT-R delay, $F(1,87) = 15.61$, $p=0.00$, $r=0.39$; HVLT-R total, $F(1,87) = 6.56$, $p=0.01$, $r=0.26$; RCFT delay, $F(1,86)=8.53$, $p=0.004$, $r=0.30$; ACE delay, $F(1,88)=12.38$, $p=0.001$, $r=0.35$ other than the PAL, $F(1, 87) = 3.903$, $p=0.051$, $r=0.21$), where differences in group means just failed to reach significance.

These findings were replicated using non-parametric group comparison methods (HVLT-R DI, $U=683.00$, $p=0.000$, one tailed , $r=-0.27$, HVLT-R delay, $U=549.50$, $p=0.000$, $r=-0.38$, HVLT-R total, $U=667.50$, $p=0.004$, $r=-0.28$;RCFT delay, $U=633.00$, $p=0.002$, $r=-0.30$, ACE delay, $U=593.50$, $p=0.000$, $r=-0.36$; PAL, $U=787.00$, $p=0.052$, $r=-0.18$.

ANVOCA revealed that the covariate FSIQ, was significantly related to performance on the delayed recall component of the ACE $F(1, 54)=13.12$, $p=0.001$; the total number of words recalled across the three learning trial of the HVLT-R $F(1,54)=5.286$, $p=0.025$ and the number of words from the HVLT-R recalled following a delay $F(1,54)=6.03$, $p=0.017$ but none of the remaining memory measures ($p>0.05$ in all cases). After controlling for group differences in FSIQ, the mean performances of the early AD group were significantly lower than that of the control group on all 6 of the memory measures (HVLT-R DI control mean(SD)=9.54(0.38), AD mean(SD)=4.56(0.65), [$F(1, 54) = 41.52$, $p = 0.000$, $r=0.66$; HVLT-R delayed recall control mean(SD)=7.19(0.46), AD mean(SD)=2.21(0.80), [$F(1,54)=27.34$, $p=0.000$, $r=0.58$;HVLT-R total recall control mean(SD)=21.88(0.87), AD mean(SD)=14.67(1.45), [$F(1,54)=16.45$, $p=0.000$, $r=0.48$; ACE delayed recall control mean(SD)=5.55(0.26), AD mean(SD)=1.98(0.46), [$F(1,54)=43.53$, $p=0.000$, $r=0.66$; PAL 6 box error score control mean(SD) 11.93(1.71), AD mean(SD)=38.60(2.96), [$F(1,54)=57.89$, $p=0.000$, $r=0.72$; Rey Figure delayed recall control mean(SD)=15.74(0.98), AD mean(SD)=4.05(1.70), [$F(1,54)=33.63$, $p=0.000$, $r=0.62$.

These findings were replicated using non-parametric group comparison methods with the following results: HVLT-R DI control median (range)=10.00(3-12), AD median (range) =5(-2 - -11), $U=69.50$, $p=0.000$, $r=-0.66$; HVLT-R delayed recall control median(range) =8.00(0-12) , AD median(range) =0.00(0-10), $U=80.00$, $p=0.000$, $r=-0.64$; HVLT-R total recall control median(range) =21.50(9-24), AD median(range) =14(4-22), $U=100.00$,

p=0.000, r=-0.60; ACE delayed recall control median(range)=6.00(0-7), AD median(range)=0.00(0-7), U=67.00, p=0.000, r=-0.69; PAL 6 box error score control median(range)=9.00(0-50), AD median(range)=42.50(13-50), U=39.50, p=0.000, r=-0.68; Rey Figure delayed recall control median(range)=16.00(0-31), AD median(range)=0.00(0-14), U=53.00, p=0.000, r=-0.70.

AUC's were universally high for the AD vs. combined control group comparison: HVLT DI = 0.92, HVLT delayed recall = 0.90, HVLT total recall across learning trials = 0.88, ACE delayed recall = 0.92, PAL 6 box error = 0.94 and Rey delayed recall = 0.94 and moderate for the MCI vs. combined control group comparison (see Table 4.5); 0.64, 0.71, 0.66, 0.72, 0.60, 0.67 respectively.

At a pre-set specificity level of 0.8, the sensitivities of cued and free recall measures to AD ranged from 0.73 for the HVLT total recall across the learning trials to 0.94 for the total number of errors made at the 6 box level of PAL. The sensitivities of the free recall tasks to MCI were a little higher than that of the cued recall measures, but remained low (see Table 4.5). At a pre-set sensitivity level of 0.8, the specificity values associated with the free recall measures for both group comparisons were consistently a little higher than those associated with cued recall measures, but remained low for the MCI comparison.

Percentage's of MCI, healthy and depressive elderly participants with overlapping scores ranged between 0% for the PAL 6 box error and ACE delayed recall scores to a maximum of 4.5% for the total number of words recalled across the three learning trials of the HVLT-R. The percentage of early AD, CT and Depressive participants with overlapping scores on the memory measures were as follows: HVLT-DI=19.04%, HVLT delayed recall=7.94%, HVLT total recall across three learning trials=33.33%, PAL 6 box errors=41.67%, ACE delayed recall=0%, Rey complex figure test delayed recall=39.06%. Dot plots of individual participant data points on each of the episodic memory measures are included in section 8.3 of the Appendix.

4.2.6 Discussion

There are a number of levels on which episodic memory paradigms can vary. Their application to the assessment of early and pre-clinical AD is, however, frequently undertaken without knowledge or consideration of how this may influence the diagnostic performance of the memory test. In this study we have compared the discriminative abilities of 6 well known episodic memory measures among groups of early AD, MCI, healthy and depressive elderly with a view to 1) testing the robustness of previous findings of superior diagnostic capacity of PAL over more traditional paper and pencil episodic memory tasks 2) demonstrating superior sensitivity of free as opposed to cued recall tasks to memory impairment in MCI and AD and 3) demonstrating superior specificity of cued (as compared to free) recall measures against normal aging and depressive symptoms among the elderly.

From the AUC's reported in the first column of Table 4.5, it is apparent that each of the 6 memory measures discriminated early AD patients from healthy and depressive elderly controls with high levels of overall accuracy. Whilst the PAL and the delayed recall component of the RCFT were associated with the highest levels of diagnostic accuracy (94%), several other of the more traditional paper and pencil style episodic memory tasks, including the HVLT-R, and the delayed recall component from the ACE, achieved similar levels of overall diagnostic accuracy.

In recognition of the necessity to achieve high levels of both sensitivity and specificity in clinical practice, these parameters were consecutively pre-set to a minimum of 80% and the resulting highest combinations of the two were determined for each memory measure. At a pre-set minimum specificity value of 80%, the PAL was the most sensitive of the episodic memory measures to early AD, and was also associated with the largest percentage (i.e. 41%) of non-overlapping scores across early AD, normal and depressive elderly participant groups. On closer examination of individual scores, (see figure 8.1 Appendix C) it is apparent that only 2 of the 16 early AD patients obtained error scores of less than 34 at the 6 box level, the removal of which would result in a considerably higher percentage, (i.e. 94 %) non-overlapping participant scores. The latter is similar to the non-overlapping score percentage of 93% reported by Swainson et al. (Swainson et al. 2001) in a slightly larger sample of 26 mild AD patients.

However, the PAL lacked sensitivity to MCI with the difference in the mean number of errors made at the 6 box level failing to distinguish these participants from a combined group of depressive and healthy elderly controls. As a result, the overall diagnostic accuracy of the PAL was lower than for measures of delayed free recall. Swainson et al.(Swainson et al. 2001) similarly failed to differentiate between groups of Questionable Dementia sufferers and combined healthy and depressive elderly control groups using the PAL and Fowler et al. (Fowler et al. 1995) reported no difference in the PAL performance of QD participants and healthy elderly controls at their initial presentation, although differences in performance emerged with time and all of the QD went on to receive a clinical diagnosis of AD.

By contrast, other authors have reported impaired performance on this measure in association with MCI as defined by Petersen (Ahmed et al. 2008b;Alladi et al. 2006;Lonie et al. 2008). It is possible that these seemingly contradictory findings are attributable to the use of healthy age matched comparison groups as opposed to mixed depressive and healthy control groups and/or the heterogeneous nature of MCI itself, with scores of future non-converters counterbalancing those of future dementia converters. For these reasons it will be of interest to determine the discriminative capacity of the PAL once our MCI participants have been sub grouped in accordance with their long term diagnostic outcome.

Whilst we failed to find evidence for greater overall diagnostic accuracy of free recall measures in differentially diagnosing early AD, the AUC's of the two delayed verbal free recall measures were highest in differentiating MCI from the combined controls. Furthermore, although none of the memory measures yielded adequately high i.e. >80% combinations of both sensitivity and specificity to MCI relative to the combined depressive and healthy elderly control groups, the sensitivities of the free recall measures were higher than those of the cued recall memory measures (at pre-set minimum 80% specificity levels). This was also reflected in the highest AUC values observed for the delayed recall components of the HVLT-R and the ACE. A similar pattern of findings for specificity values were observed when sensitivity levels were pre-set at a minimum of 80%. Together these findings support the use of free recall measures in the assessment of MCI and imply they can be applied without risk of loss of specificity. More importantly, clinicians should be aware of the trade off between achieving high levels of sensitivity to memory impairment in MCI and specificity against normal aging and depression when assessing these patient groups.

Thus, the free recall measures employed within the present study appear equally as specific as cued recall measures in the differential diagnosis of early AD. Some gain in sensitivity is seen using free recall memory measures in assessing MCI and this finding is in keeping with the larger effect sizes that have been reported in association with measures of free, relative to cued recall in MCI samples (Alladi et al. 2006;Lonie et al. 2008). However no single memory measure provides acceptable combined levels of sensitivity and specificity and sensitivity levels remain universally low at acceptable (i.e. 80% or above) levels of specificity.

In this study, patients who fulfilled Petersen's (Petersen et al. 1999) MCI criteria performed significantly less well than a group of combined depressive and healthy elderly controls on a cued measure of delayed verbal recognition but not on a cued measure of visuo-spatial learning ability. The differences in findings across the two measures are conceivably accountable for by other (non-cued vs. free recall) levels of task variability such as the length of delay or the potential for interference from other tasks, and it is likely that these factors, together with variability in the general level of cognitive functioning and pathological make-up of MCI samples account for some of the cross study inconsistency in reports of performance of MCI subjects on cued recall measures.

With sensitivity levels pre-set at a minimum of 80%, there was no evidence for superior specificity of cued, over free recall tasks among the selected 6 memory measures. As noted above, however, the least degree of overlap in scores was observed for the PAL of the CANTAB battery in differentiating AD from combined controls, and this may be attributable to any number of factors including the associative, cross modal or cued aspects of the task. These findings have relevance for clinical practice where cut off points are routinely used to classify patients on an individual basis and imply that sound levels of specificity are attainable in relation to mild-moderate levels of depressive symptoms among the elderly in assessing for AD and MCI using free recall memory tasks.

4.2.7 Conclusion

To summarise, both cued and free recall memory paradigms differentiate early AD from normal aging and cognitive difficulties associated with depressive symptoms with high levels of sensitivity and specificity. Whilst the PAL was the most sensitive of the 6 episodic

memory measures to AD, and was associated with the least degree of overlapping scores, several other more traditional paper and pencil measures achieved similar levels of overall diagnostic accuracy in classifying early AD patients, normal and depressive elderly. At a designated 'acceptable' minimum level of sensitivity (i.e. 0.8), the cued recall measures employed in the present study were not associated with higher levels of specificity in discriminating early AD or MCI participants from depressive and healthy elderly controls. None of the memory measures were adequately sensitive and specific to MCI, although free recall measures showed the best combination of these properties. These findings support the use of free recall measures in the early differential diagnoses of AD and MCI and imply that the use of the PAL subtest from the CANTAB battery may be of particular value in safeguarding against false positive early AD diagnoses. The results also highlight the limitations of existing episodic memory measures in differentially diagnosing MCI.

4.3 A comparison of object and face naming tasks in the differential diagnosis of AD, MCI, and the healthy and depressive elderly.

4.3.1 Abstract

Background

Although memory impairment is the defining characteristic of aMCI, there is evidence to suggest that impairments are also present in other cognitive domains and that these may be of importance in reaching an early diagnosis of AD. Impairments of naming ability are frequently reported in association with aMCI, and are also known to co-exist with episodic memory impairment in the very early stages of AD. Object naming ability has traditionally been assessed within English speaking cultures using one of two well-established neuropsychological measures i.e. the Boston Naming Test (BNT) or the Graded Naming Test (GNT). It is unclear whether or not these measures perform comparably in a differential diagnostic sense. Furthermore some recent findings suggest that tests of person specific knowledge may be more sensitive to early semantic memory impairment in pre-clinical and early AD than these more commonly applied object naming tasks.

Aims

In this study we investigate the relative diagnostic accuracy of 3 naming tasks (2 object and 1 famous face) in classifying groups of early AD, aMCI, and combined healthy and depressive elderly controls. A decision to combine the healthy and depressive elderly control groups was made in accordance with the primary focus of the present study, in comparing 1) the sensitivities of face and object naming tasks to early AD pathology and 2) the abilities of the two most commonly used object naming tasks to differentatie between word finding difficulties due to early AD pathology as opposed to other conditions commonly encountered in clinical practice (i.e. normal aging and depression). We sought to test the robustness of recent findings of superior and differential sensitivity of face as compared to object naming tasks to aMCI and early AD, as well as to establish the specificity of famous face naming against healthy and depressive elderly controls. We further sought to compare the performances of two well-established object naming tasks (BNT & GNT) in differentiating patients with early AD and MCI from the healthy and depressive elderly.

Methods

The discriminative capacities of 3 naming tasks (2 object and 1 famous face) for groups of 46 participants with MCI and 20 with early AD, relative to a combined control group comprising 20 elderly out patients with depressive symptoms, and 24 healthy age matched elderly, were determined by way of group mean comparisons. Where significant group differences were present, AUC, sensitivity and specificity values were computed by way of ROC analyses to facilitate comparison of their diagnostic utilities. An object vs. face naming difference score was derived for each participant by subtracting the percentage of famous faces correctly named on the GFT from the average percentage of objects correctly named on the BNT and GNT. Group differences in the magnitude of mean object-face naming discrepancy scores were analysed to determine the specificity of a GFT < GNT pattern of performance to AD and MCI participant groups.

Results

The early AD participant group performed significantly less well than the combined control group on all three naming measures. The GFNT was the only measure of the three that discriminated MCI participants from a combined healthy and depressive elderly control group. The largest AUC's were reported in association with the GFNT (AD vs. CT= 0.93; MCI vs. CT=0.66). Whilst the magnitude of GNT-GFT difference scores did not differ significantly across the three participant groups, as a combined group, AD and MCI participants performed proportionately worse on the GFT relative to the GNT, than the combined (i.e. depressive and healthy) controls.

Conclusion

The GFNT appears to be a particularly sensitive measure of naming ability in the early stages of AD and in patients who fulfil Petersen's criteria for MCI. Preliminary findings further suggest that poor performance on this measure is relatively specific against the cognitive effects of normal ageing and depressive symptoms. The GFNT may therefore be more suited for use in the early differential diagnosis of AD than traditional object naming tasks (which show similar and comparatively lesser levels of discriminative capacity). The level of impairment on famous face naming task, in relation to object naming ability, was greater for combined groups of MCI and early AD patients than for healthy and depressive elderly controls, providing some tentative evidence for a selective vulnerability of face over object naming in AD. Further longitudinal analysis will allow for determination of the relative prognostic contributions of these 3 naming measures within our MCI sample.

4.3.2 Introduction

Although memory impairment is the defining characteristic of aMCI, there is evidence to suggest that impairments are also present in non-memory domains (Alladi et al. 2006;Kramer et al. 2006;Lonie et al. 2008;Lonie et al. 2009b;Lonie et al. 2009a) and that these may be of importance in reaching an early diagnosis of AD. Naming difficulty (anomia) is one of the most commonly reported cognitive complaints of a non-amnestic nature in early AD and aMCI. Naming impairments have traditionally been documented clinically via performance on object naming tasks such as the Graded Naming Test (GNT; (McKenna and Warrington 1980;Warrington 1997) and the Boston Naming Test (BNT; (Kaplan et al. 1983).

In AD, impairment on naming measures is attributed to deficient activation at the conceptual level of the semantic lexicon (conceptual-semantic) or at the final stages of name retrieval (lexical-semantic) or indeed both (Ahmed et al. 2008a;Grabowski 2008). The neuroanatomical underpinnings of object and face naming, as determined by imaging studies, are known to overlap with the spread of pathology from the medial to the more lateral aspects of the temporal lobe that characterises the neocortical stage of AD (Braak and Braak 1991), thus providing a sound theoretical underpinning to naming impairments in early and pre-clinical AD (Grabowski 2008). Alternatively, if episodic and semantic memory were both equally dependent on the integrity of the medial temporal lobe structures, as proposed by Squire (Squire 1992), then pathology in this region alone could conceivably give rise to semantic memory impairments in early and pre-clinical AD.

Not surprisingly then (if we are to assume that at least a proportion of those with aMCI will progress to AD) object naming impairments have also been reported in association with aMCI (Ahmed et al. 2008a;Alladi et al. 2006;Dudas et al. 2005b;Loewenstein et al. 2006a;Lonie et al. 2008;Thompson et al. 2002). More recently, however, the focus has shifted toward naming tasks comprising unique entities, such as famous faces, buildings or events, where each item is associated with a single representation within the semantic memory system and where attributes of that single representation are arbitrary (Ahmed et al.

2008a;Clague et al. 2005;Dudas et al. 2005b;Estevez-Gonzalez et al. 2004;Joubert et al. 2008;Vogel et al. 2005).

There is some preliminary evidence to suggest that naming tasks of this nature are differentially impaired in aMCI (Ahmed et al. 2008a;Clague et al. 2005;Joubert et al. 2008) and early AD (Thompson et al. 2002). The greater sensitivity of face over object naming tasks has been attributed to the differing manner in which faces and objects are represented within the semantic memory system (Ahmed et al. 2008a;Grabowski 2008;Joubert et al. 2008), (as noted above, in the case of faces, representations are singular and attributes are arbitrary) or alternatively to the existence of a unique neuroanatomical network dealing specifically with knowledge of faces (Estevez-Gonzalez et al. 2004). Recent findings of a differential impairment of famous building and events knowledge in MCI would, however, argue against the latter explanation (Ahmed et al. 2008a;Joubert et al. 2008).

It is argued that involvement of areas outwith the medial temporal lobe, in particular the temporal neocortex, as proposed by Braak (Braak and Braak 1991), may be an important indicator of likely progression toward AD (Grabowski 2008). For this reason, naming measures may prove useful in helping to differentiate MCI as it represents pre-clinical AD from other non-progressive conditions, such as depression, that also give rise to recent memory impairment.

There is some preliminary evidence to support this claim, with two studies having reported a significantly poorer mean performances in pre-clinical AD sufferers relative to age matched controls, on measures of face naming ability (Estevez-Gonzalez et al. 2004;Vogel et al. 2005). Thompson et al (Thompson et al. 2002) also reported a NPV of .94 in association with the GFNT and positive predictive value of 1 for the GNT in detecting MCI as it represents the pre-clinical phase of AD, among a very small group of aMCI patients.

Of note, however, the sensitivity of the GNT to pre-clinical AD was reportedly low (14%) in this, as well as other (Fox et al. 1998;Godbolt et al. 2004) studies (Blackwell et al. 2004). Albert et al (Albert et al. 2001) similarly failed to find any significant difference in the baseline BNT scores of QD patients who went on to develop dementia and those who did not and several (De Jager et al. 2003;Loewenstein et al. 2007b;Petersen et al. 1999) but not all

studies (Balthazar et al. 2007;Greenaway et al. 2006) have reported significantly lower BNT scores in association with MCI relative to age matched controls.

If naming tasks were to be of diagnostic and/or prognostic utility in aMCI, as has been suggested, specificity against the normal aging process, as well as a range of non-progressive psychiatric and medical conditions would need to be high. Whilst a differential GFNT deficit unique to AD (Clague et al. 2005) and MCI (Ahmed et al. 2008a;Joubert et al. 2008) has been reported, poor performance on the GFNT does not appear to be specific to AD, and has also been reported in association with other forms of dementia including vascular dementia, semantic dementia (Clague et al. 2005), and dementia with lewy bodies (Troster 2009). The specificity of face naming tasks against conditions of a non-progressive nature, such as depression, has not been examined. In the case of object naming, Swainson et al (Swainson et al. 2001) reported no difference in the performances of QD participants and a combined depressive and healthy elderly control group on the GNT, although the extent to which the QD group comprised patients who were in the pre-clinical phase of AD is unclear.

In this study we investigate the relative diagnostic accuracy of 3 naming tasks (2 object and one face) in classifying groups of early AD, MCI, and combined healthy and depressive elderly controls. Within a clinical context, neuropsychological performance is used to assist in separating out the frequently simultaneously occurring effects of aging and depression on cognitive performance, from the early or pre-clinical cognitive manifestations of AD. The depressive and healthy elderly control groups, (who are well matched in terms of age, general levels of cognitive functioning and NART FSIQ), were therefore combined in an attempt to simulate clinical practice and maximise the clinical relevance of the study findings. We sought to test the robustness of recent findings of a differential face naming deficit in MCI and early AD as well as to establish the specificity of this measure against normal aging and symptoms of depression among the elderly. We further sought to compare the performances of two well-established object naming tasks (BNT & GNT) in differentiating patients with early AD and MCI from the healthy and depressive elderly. Longitudinal follow-up would allow for later determination of the prognostic utility of these naming measures.

4.3.3 Method

Groups of healthy elderly (n=24), depressive elderly (n=20), early AD (n=20) & MCI participants (n=46) were administered three naming measures, the Graded Naming Test (GNT), the Boston Naming Test (BNT) and the Graded Faces Naming Test (GFNT), together with two measures of global cognitive functioning (MMSE & ACE) and a measure of pre-morbid intellectual functioning (NART), as part of a wider neuropsychological test battery.

Details of the method of subject recruitment, participant characteristics together with psychometric and administrative characteristics of the naming measures employed can be found within Chapter 2, Materials and Methods, sections 2.1 and 2.

4.3.3.1 Statistical Analysis

Differences in gender proportions between participant groups were examined using the chi-square test. In view of violations of the assumptions of both normality and homogeneity of variance (see below), differences in the mean age and NART FSIQ of the participant groups were analysed using the Kruskal-Wallis Test followed by Man-Whitney post hoc comparisons with the Bonferroni level of correction set at $p<0.05$ / the total number of comparisons.

Visual inspection of the data together with Kolmogorow-Smirnov tests revealed that the distributions of BNT scores for the MCI and combined depressive and healthy control groups were not normal. Distributions were also abnormal for the combined healthy and depressive control groups on the GNT and GFT; and for the MCI and combined control group FSIQ scores; and the age scores for the early AD participants. Furthermore, assumptions of homogeneity of variance were not universally met.

As the NART estimated FSIQ score was significantly lower for the early AD group relative to the combined controls, an ANCOVA was carried out with NART FSIQ as a covariate to examine group differences in scores on each of the 3 naming measures.

As the NART estimated FSIQ scores of the MCI and combined control groups were not significantly different, but assumptions of normality and homogeneity of variance were not universally met for either of the two group comparisons, the Kruskal-Wallis test was also conducted to examine group differences in mean scores on each of the 3 naming measures. This was followed up with Mann-Whitney post hoc group comparisons (where the Bonferroni level of correction was set at $p<0.05$ / the total number of comparisons (accounting for the 6 comparisons).

ROC analyses were used to compare the diagnostic accuracy of each of the three naming measures.

Difference scores were calculated for each participant by subtracting the total number of famous faces correctly named on the GFNT from the total number of objects correctly named on the GNT. Visual inspection together with Levene's and Shapiro-Wilk tests revealed that the difference scores for each of the three participant groups were normally distributed and variances were homogenous. Mean difference scores for the three participant groups were therefore examined using ANCOVA to determine whether the pattern of performance on the two types of naming measures differed as a function of participant groups. Planned contrasts were also conducted to compare the magnitude of difference in GNT and GNFT scores across control and patient groups and across the two (i.e. MCI and early AD) patient groups.

4.3.4 Results

There was no significant association between participant group and gender, Chi-squared(2) = 2.07, $p>0.05$. The mean ages of the participant groups did not differ significantly, ($H(2) = 2.40$, $p>0.05$). NART FSIQ scores differed significantly as a function of participant group, ($H(2) = 9.409$, $p<0.01$). NART estimated FSIQ scores for the MCI (median=120) and combined healthy and depressive control group (median=118) did not differ significantly ($U=924.5$, ns two-tailed, $r=-0.04$), whereas the NART estimated FSIQ score of the early AD group (median=110.5) was significantly lower than that of the combined control group (median=118.00; $U=231.00$, $p<0.01$ two-tailed, $r=-0.37$).

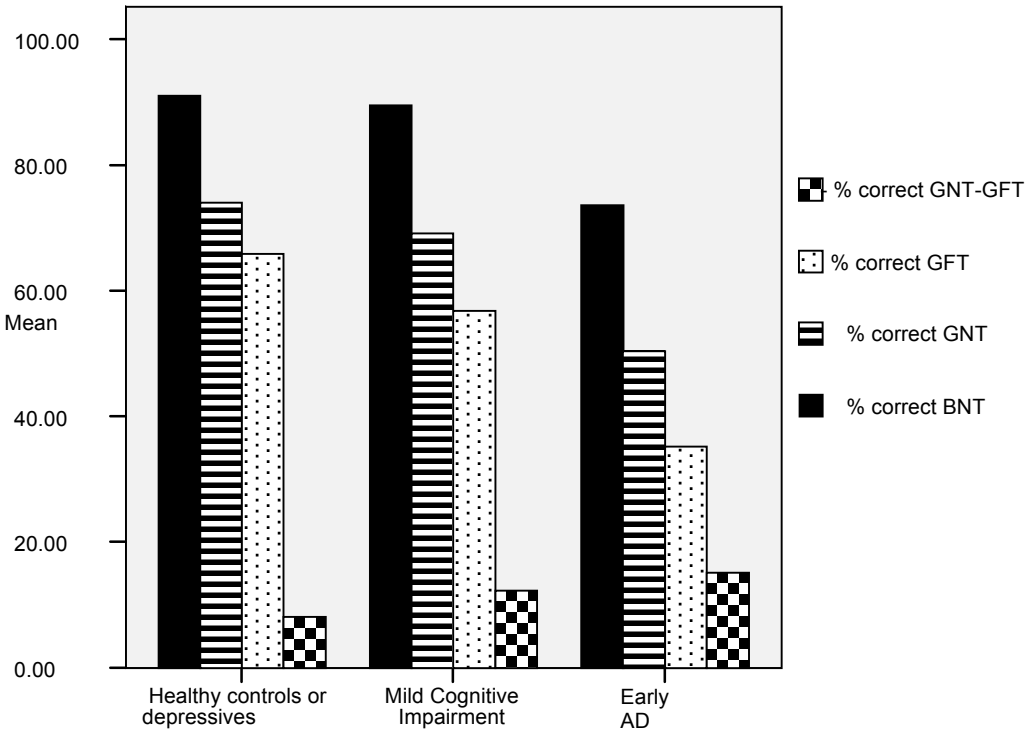
The covariate NART FSIQ, was significantly related to naming performance on the BNT $F(1,54)=16.17$, $p<0.001$, $r=0.48$; the GNT $F(1,54)=16.25$, $p<0.001$, $r=0.23$ but not the GFT $F(1,54)=1.271$, $p>0.05$, $r=0.15$. There was also a significant effect of participant group on naming ability (i.e. AD vs. CTs combined) after controlling for the effect of NART FSIQ; BNT $F(1,54)=11.226$, $p<0.01$, $r=0.41$; GNT $F(1,54)=14.26$, $p<0.001$, $r=0.45$; GFT $F(1,54)=38.65$, $p<0.001$, $r=0.64$).

Table 4.6 Demographic and neuropsychological data for groups of aMCI, Early AD and combined depressive and healthy elderly control groups.

	Combined Healthy and Depressive Controls (n=39-44) Mean(SD) Median(range)	MCI (n=43-46) Mean(SD) Median(range)	Early AD (n=18-20) Mean(SD) Median(range)	Combined CT vs. Early AD Effect sizes (r)	Combined CT vs. MCI Effect sizes (r)
Age	72.25(1.12) 72.00(59-85)	73.89(0.95)ns 74.50(58-85)	75.30(1.46)ns 78.00(64-86)	r=-0.17	r=-0.12
Sex M:F	30:14	27:19	10:10	_	_
FSIQ	117.56(0.08) 118.00(100-126)	116.78(1.13)ns 120.00(92-126)	112.20(1.57)** 110.00(99-124)	r=-0.37	r=-0.04
MMSE	28.80(0.98) 29.00(25-30)	28.35(1.48)ns 29.00(24-30)	24.25(3.28)*** 24.50(17-29)	r=-0.68	r=-0.13
ACE	92.18(5.56) 94.00(78-99)	89.46(5.52)* 89.00(77-99)	75.95(9.02)*** 76.50(58-91)	r=-0.71	r=-0.27
BNT	54.98(5.83) 57.00(29-60)	53.59(5.32)ns 55.00(39-60)	45.05(10.20)*** 49.00(19-56)	r=-0.57	r=-0.19
GNT	22.21(4.64) 23.00(5-29)	20.74(4.12)ns 22.00(10-27)	15.11(5.12)*** 16.00(7-24)	r=-0.58	r=-0.2
GFNT	19.93(4.34) 21.00(9-27)	16.91(4.96)** 18.00(5-27)	10.50(4.05)*** 11.00(2-18)	r=-0.70	r=-0.32
GNT - GFNT	2.44 (4.98)	3.70 (5.02)	4.56 (2.68)	_	_

FSIQ = Nart based Full Scale Intelligence Quotient; MMSE=Mini Mental State Examination; ACE=Addenbrookes Cognitive Examination; BNT=Boston Naming Test; GNT=Graded Naming Test; GFNT=Graded Faces Naming Test; MCI=Mild Cognitive Impairment; AD=Alzheimer's disease; *=p<0.05; **=p<0.01; ***=p<0.001; ns=non significant following Bonferroni adjustment for multiple comparisons

Figure 4.4 Mean percentages of correctly named items on the object and face naming tasks as a function of participant group

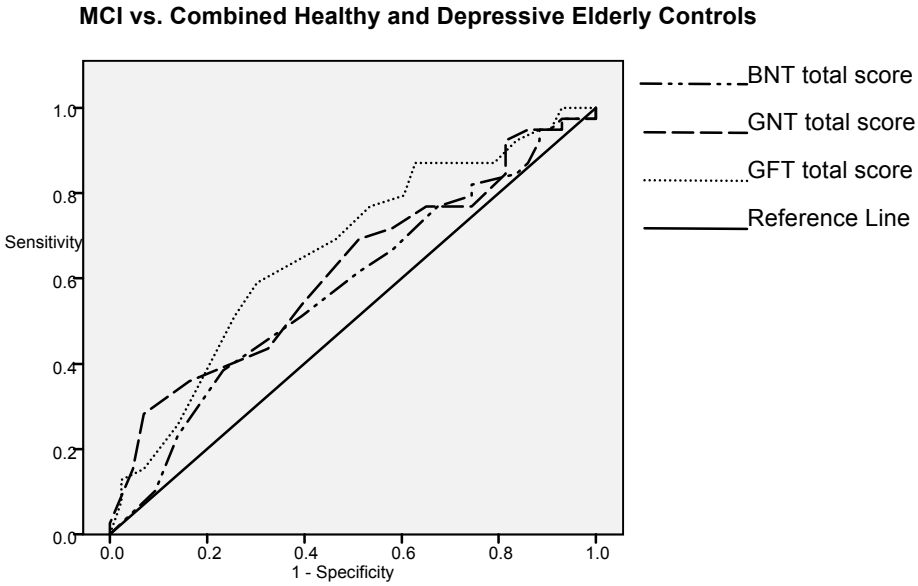


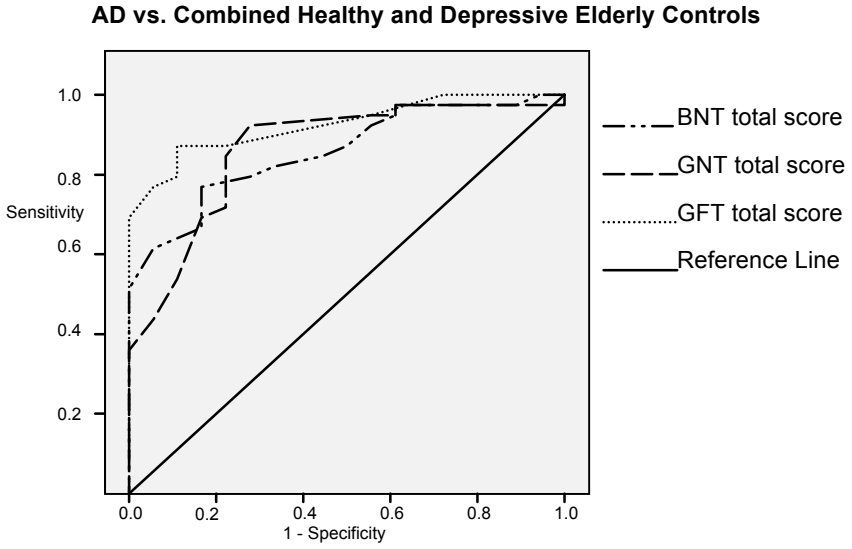
These findings were replicated using non-parametric analysis methods with the critical value for significance set at $p<0.0083$ on account of the 6 comparisons that were conducted (see below) (GNT, $U=96.00$, $p<0.001$, $r=-0.58$; BNT, $U=124.00$, $p<0.001$, $r=-0.57$; GFNT, $U=56.50$, $p<0.001$, $r=-0.70$).

The mean number of faces correctly named by the combined control group was significantly higher than for the MCI participant group ($U=624.50$, $p<0.01$, $r=-0.32$). Group differences otherwise failed to reach significance following Bonferroni corrections for multiple comparisons (with $p<0.0083$ set as the critical value for determining significance; BNT, $U=785.50$, $p<0.05$, $r=-0.19$; GNT, $U=643.50$, $p<0.05$, $r=-0.2$).

The mean GNT-GFT discrepancy scores for the 3 participant groups did not differ significantly $F(2,97) = 1.46$, $p>0.05$, $r=0.17$ however, planned contrasts revealed that the magnitude of the GNT-GFT score discrepancy was significantly greater for the combined patient (i.e. AD + MCI) than control (i.e. Depressive + Healthy elderly) group, $t(97)=1.70$, $p<0.05$, $r=0.17$. No difference in the magnitude of GNT-GFT discrepancy was observed between the early AD and MCI patient groups, $t(97)=0.65$, $p>0.05$, $r=.0.07$

Figure 4.5 ROC curves comparing the classification accuracies of the GNT, BNT and GFNT among MCI and combined healthy and depressive elderly control participants and early AD and combined control groups.





AUC's for the AD vs. CT comparison were universally high (BNT=0.86; GNT=0.86; GFNT=0.93) but poor for the MCI vs. CT comparison (BNT=0.58; GNT=0.62; GFNT=0.66). For both participant group comparisons, the GFNT was associated with the highest overall discriminative value.

Figure 4.6 Dot plot of individual participant scores on the GFNT in accordance with participant group

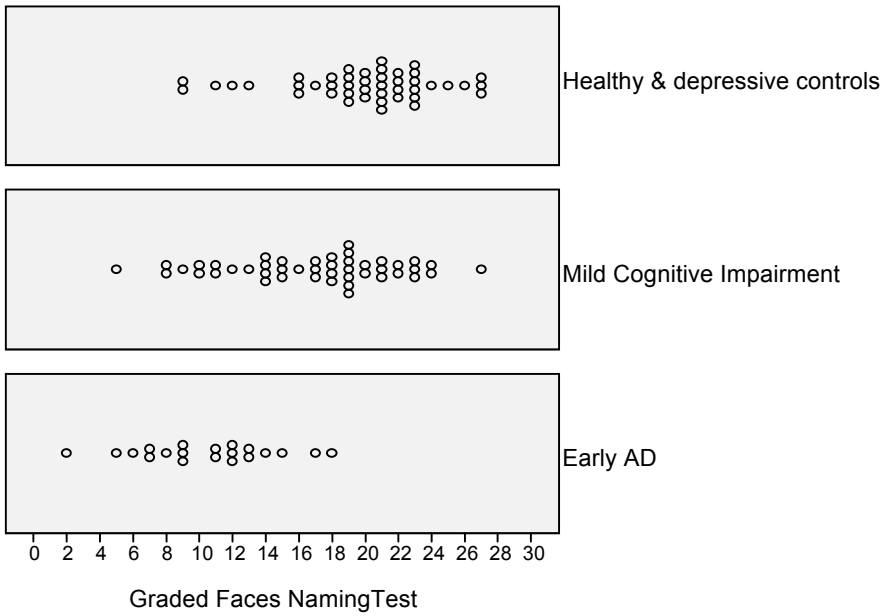


Figure 4.7 Number of MCI participants performing 1.5 SD below healthy elderly control participants on the famous faces naming test

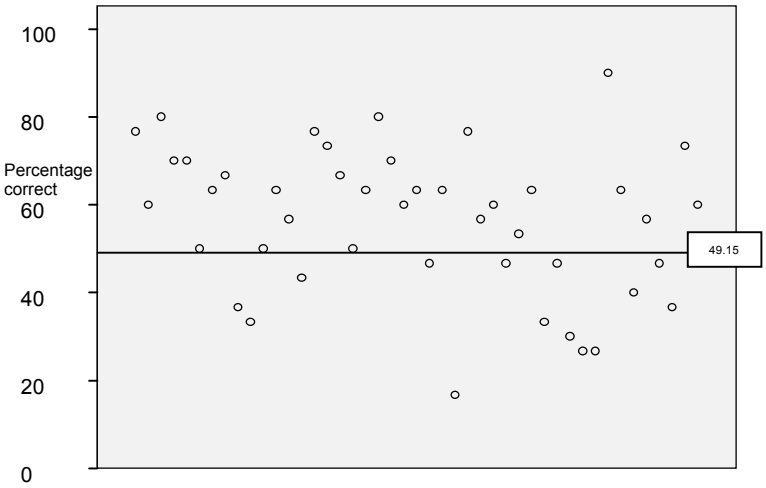
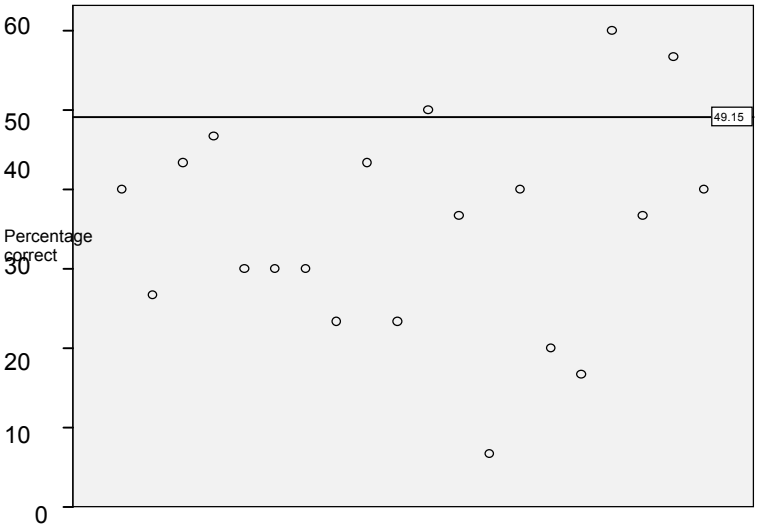


Figure 4.8 Number of early AD patients performing 1.5sd below the healthy elderly control participant group mean on the GFNT



4.3.5 Discussion

The high (i.e. AUC 0.86 – 0.93) levels of overall diagnostic accuracy reported in this study indicate that naming tasks are useful measures in the differential diagnosis of early AD.

Levels of diagnostic accuracy for the object naming tasks (AUC=0.86) were comparable to those seen in association with verbal learning tasks (i.e. HVLT-R total recall across three trials AUC = 0.88), and the choice of object naming task (i.e. BNT vs. GNT) appeared to have no effect on overall diagnostic accuracy. Of greater note, the diagnostic accuracy of the famous face naming task was as high as that reported in association with a computerised measure of visuospatial associative learning ability (CANTAB PAL), suggesting that face naming tasks may be of equal importance as episodic memory measures in the neuropsychological evaluation of early AD.

Whilst the findings are in keeping with the broader notion of failure(s) at one or more level(s) of the lexico-semantic system representing an early feature of AD, they do not address the precise level at which breakdown is occurring, or to what extent that breakdown is attributable to underlying temporal neocortical Alzheimer pathology. It is nonetheless conceivable that the very high level of overall diagnostic accuracy observed in association with the GFNT is attributable to 1) the specificity of naming tasks to AD (against depression and normal ageing) as a proxy for pathological change in temporal neocortex, and/or 2) the preferential vulnerability of naming items with single arbitrary representations, within the semantic memory system seen in association with this pathological change.

In the present study, object (i.e. GNT) and face naming (i.e. GFNT) tasks were not matched for difficulty, and it can be seen (figure 4.4) that each of the three participant groups obtained lower mean total scores on the famous face naming as compared to the graded naming task. This finding is in keeping with that of Ahmed et al (Ahmed et al. 2008a) where control patients displayed a similar pattern of performance across naming tasks i.e. GNT>GFT to that of MCI participants. Although the relative differences in mean scores on the GFT and GFNT (i.e. GNT-GFNT) did not differ significantly between our three participant groups, the magnitude of difference (i.e. an advantage of GNT over GFNT) was significantly greater for the combined patient (MCI or AD) vs. control groups. A similar observation was again made by Ahmed et al (Ahmed et al. 2008a) where differential impairment was reportedly significantly greater for the GFT than the GNT among MCI participants relative to healthy elderly controls. Furthermore, with object and famous face naming tasks matched for levels of difficulty, Clague et al (Clague et al. 2005) reported advantages in performance on the GNT relative to GFNT for QD and AD patients but not controls.

Taken together, these findings provide some support for the specificity of selective vulnerability of famous face naming in AD, suggesting that unique exemplars may indeed be significantly more difficult to retrieve than common exemplars as a result of the AD disease process. It should be noted, however, that the GFNT used by Clauge et al (Clague et al. 2005), appears to differ from that utilised in the present study and that of Ahmed (Ahmed et al. 2008a), both of which comprised 30 as opposed to 32 famous faces, complicating study comparison.

In the present study the GFNT was the only naming measure to differentiate between MCI and control participants. Calgue et al (Clague et al. 2005) similarly reported a significant difference in the mean performances of QD sufferers and healthy elderly controls on the GFNT but not the GNT. In a study by Vogel et al (Vogel et al. 2005) measures of famous face naming ability and category fluency, but not of object naming ability, differentiated 22 pre-dementia AD sufferers from 58 matched controls. Thompson et al (Thompson et al. 2002) documented differences in the baseline performances of MCI patients who went on to develop dementia relative to those who do not, on the GFNT but not the GNT, whilst others have failed to demonstrate group differences in the performance of participants in the pre-clinical phase of AD and age matched controls on the GNT (Fox et al. 1998;Godbolt et al. 2004). Our findings therefore add to a growing evidence base suggesting famous face naming measures may be more sensitive than object naming measures to the semantic impairment that characterises pre-clinical and early stage AD.

The ability of the naming tasks to differentiate MCI sufferers from elderly control patients in the present study, was modest, at best, and although greatest for the famous face naming task, high levels (i.e. 96%) of overlapping scores among the MCI and combined control groups were apparent on this naming task. A cut off score of 16/30 correctly classified 88.6% of controls whilst essentially dividing the MCI participant group in half, with 18/44 obtaining a score of less than 16/30 and the remaining 27/44 obtaining scores above this level. It is plausible that the lower overall discriminative power of the naming tasks in relation to MCI reflects the heterogeneous nature of this participant group, some but not all of whom will likely progress towards a clinical diagnosis of dementia over time. Longitudinal follow-up of the MCI cohort will enable us to determine the prognostic significance of this split.

Higher levels of classification accuracy (i.e. 100% CT; 78% MCI) have been reported in association with this measure (Ahmed et al. 2008a) although it is noted that the control group comprised healthy non-depressed elderly as opposed to the combination of healthy and depressive elderly used in the present study. Furthermore, the mean general level of functioning of the MCI participant group as indicated by the ACE in the latter study (i.e. 82/100, the suggested lower boundary cut off point for dementia (Mathuranath et al. 2000) was, however, somewhat lower than in the present study (89/100). Together with the relatively low mean age (69.5 years) of MCI participants in the study by Ahmed (Ahmed et al. 2008a), this raises the possibility that at least some of the participants comprising this MCI group were in fact already in the early stages of AD. Without knowledge of the fate of these MCI participants, it is difficult to extrapolate fully, the practical or theoretical meaning of the high classification rates reported.

One limitation of the present line of research relates to the culturally bound nature of famous face, and to a lesser extent, object naming tasks. As a result, the application of naming tasks within a study context varies widely, with very few studies to date having employed identical measures to address a common research question. Thus whilst there is mounting evidence to suggest that famous face naming tasks are an important component of cognitive assessment in early AD, it will likely be necessary to replicate existing findings, that relate to a British patient population, among a range of additional face and object naming measures of relevance to patients with alternative cultural backgrounds.

4.3.6 Conclusion

Naming tasks appear to be important measures of cognitive function in the early and differential diagnosis of AD. Measures of famous face naming ability are particularly sensitive to lexico-semantic deficits in the early stages of AD and in patients who fulfil Petersen's criteria for aMCI. Preliminary findings further suggest that poor performance on this measure is relatively specific against the cognitive effects of normal ageing and depressive symptoms. The GFNT may therefore be more suited for use in the early differential diagnosis of AD than traditional object naming tasks (which show similar and comparatively lesser levels of discriminative capacity). The level of impairment on a famous face naming task, in relation to object naming ability, was greater for combined

groups of MCI and early AD patients than for healthy and depressive elderly controls in the present study, providing some evidence for a selective vulnerability of face over object naming in AD. Further longitudinal analysis will allow for determination of the relative prognostic contributions of these 3 naming measures within our MCI sample.

4.4 Fluency discrepancy scores in the early and differential diagnosis of MCI and AD

(Published in part in The Journal of Neuropsychology. 2009; Vol 3. Part 1; 79 - 92)

4.4.1 Abstract

Background

Episodic memory is compromised in amnestic Mild Cognitive Impairment (aMCI), but lesser deficits in other cognitive domains are also commonly observed and may be helpful in identifying this group.

Aims

The relative difference in performance on lexical and semantic fluency tasks may be a sensitive and specific measure in aMCI and early Alzheimer's disease (AD).

Method

We compared four groups of participants, 35 early AD, 47 aMCI, 24 healthy controls, and 18 depressive out-patient controls, on semantic and lexical fluency as well as other neuropsychological tests. The elderly depressive group was examined independently of the healthy elderly control group in view of the discrepant meta-analytic findings for fluency performance patterns among patients with depression.

Results

Early AD and aMCI patients showed a distinct pattern of semantic impairment in the two fluency measures compared with the healthy and depressive controls.

Conclusion

The findings implicate early failure of the semantic memory system in aMCI and AD and suggest that consideration of the discrepancy in performance on semantic and lexical fluency measures may help in the early identification of AD.

4.4.2 Introduction

A recent meta-analysis demonstrated that Alzheimer's disease (AD) patients are significantly more impaired on measures of semantic than lexical fluency (Henry et al. 2004). Semantic memory impairment in AD is specific in that it is not associated with more generalised deficits in verbal intelligence or psychomotor speed. This pattern of impairment in verbal fluency measures is qualitatively distinct from the usual finding of superior semantic fluency in healthy controls (Spreen and Strauss 1998). As the category task is thought to rely more heavily on access to representations of semantic concepts than the letter task, the pattern of findings in AD is presumed to reflect degradation in the structure, content, or activation of the semantic memory system (Auriacombe et al. 2006;Jefferies and Lambon Ralph 2006;Jones et al. 2006).

Patients in the preclinical stages of AD exhibit a semantic fluency deficit, at a time when lexical fluency performance remains intact (Auriacombe et al. 2006;Beatty et al. 2002;Swainson et al. 2001). Similarly, patients who fulfil criteria for amnestic mild cognitive impairment (aMCI; (Grundman et al. 2004;Petersen et al. 1999) generate fewer words from a specified category than do age-matched controls. In contrast, they perform at normal levels on lexical fluency tasks (Alladi et al. 2006;Dudas et al. 2005b;Lonie et al. 2008;Murphy et al. 2006).

A pattern of worse semantic than lexical fluency has also been reported in patients with depression (Christensen et al. 1997;Zakzanis et al. 1998), although a more recent review suggests equal impairment of performance across the two fluency tasks that is thought to reflect a generalised reduction in processing speed (Henry and Crawford 2005).

If semantic and lexical fluency discrepancy (FD) scores are abnormal in some patients with aMCI, their magnitude and direction may prove helpful in diagnosis or prognosis. As an individually calibrated marker of performance, the direction of the discrepancy would have the advantage of being free from the need for age, gender, education, or IQ-dependent cut-off values, which require a sizable normative comparison group.

Furthermore, if depressive symptoms were associated with equivalent reductions in lexical and semantic task performance (as the processing speed account would predict), then FD scores might also be of value in distinguishing between depressive and early Alzheimer related cognitive impairment. To our knowledge, no study has examined semantic and phonemic fluency in aMCI compared with healthy controls, early AD patients and patients with depressive symptoms.

Here, we aimed to (1) replicate the finding that patients with aMCI demonstrate a pattern of fluency performance similar to that observed in patients with early AD and (2) to investigate whether their FD performance is abnormal when compared to healthy controls and out-patients with depressive symptoms.

4.4.3 Method

4.4.3.1 Patient groups

We examined 47 patients with aMCI, 18 out-patients with depressive symptoms, 24 healthy control patients, and 35 patients with mild AD. The 47 aMCI patients were recruited over a 3-year period (September 2003–September 2006) from tertiary referrals to our neuropsychological assessment service and met criteria for aMCI (Petersen et al. 1999). All aMCI patients underwent comprehensive neuropsychological and psychiatric evaluation, medical screening (including blood screen) as well as neuroimaging (CT and/or MRI or SPECT). Thirty-five patients with a NINCDS/ADRDA diagnosis (McKhann et al. 1984) of probable AD were identified as part of a clinical audit of service referral numbers and diagnoses. All early AD patients scored 17/30 or above on the mini mental state examination (MMSE; (Folstein et al. 1975) and 58/100 or above on the more comprehensive Addenbrookes Cognitive Examination (ACE; (Mathuranath et al. 2000) indicating a relatively mild disease severity. As healthy controls we recruited 24 spouses or carers of patients who had attended the neuropsychological assessment service. Participants with potentially confounding co-morbid medical, psychiatric, or neurological conditions (i.e. stroke or cerebrovascular disease, head injury, alcoholism, schizophrenia, etc.) were excluded.

In an attempt to ensure a control group of similar illness severity and general level of functioning to the aMCI patients, we recruited 18 out-patients with depressive symptoms from hospital out-patient clinics and day hospitals, who were receiving the same level of out-patient care as our aMCI group. All participants in this group presented with depressive symptoms, thought not to be primarily organic in nature, yet known to have effects on cognitive functioning both during illness and after recovery (Herrmann et al. 2007). We considered matching of illness severity to be important, as in clinical practice the differentiation of severe depression and early dementia states is less problematic than separating the sequelae of the milder forms of these disorders. Furthermore, fluency measures have been shown to be sensitive to even mild depressive symptoms (Ravdin et al. 2003). We included patients with a variety of disorders, as the type of depression does not appear to influence the magnitude of cognitive deficit (Christensen et al. 1997). Patients with any co-morbid medical, neurological, or psychiatric condition with the potential to affect cognitive function were excluded. The mean geriatric depression scale score (Yesavage et al. 1983) for this group was 14.3 (SD = 7.79) indicating mild, yet clinically significant levels of depressive symptoms.

4.4.3.2 Neuropsychological measures

Details of the method of subject recruitment, participant characteristics together with psychometric and administrative characteristics of the neuropsychological measures employed can also be found within Chapter 2, Materials and Methods, sections 2.1 and 2.

4.4.3.2.1 Verbal fluency

All patients were given two versions of the verbal fluency task. In the lexical version, patients were asked to generate as many words as possible within 1 minute beginning with the letter ‘P’. The letter ‘P’ was chosen in place of the more widely known ‘F’, ‘A’, ‘S’ as it forms part of the ACE (Mathuranath et al. 2000). In the second task, patients were asked to provide as many animal names as possible in 1 minute, as a subtest of the ACE. Patients’ scores were z-transformed using control means and standard deviations to be able to compare lexical and semantic task performance.

4.4.3.2.2 Episodic memory

All groups were administered the Hopkins verbal learning test – revised (HVLT-R; (Brandt 1991) where measures of interest included; the total number of words recalled across three registration trials (max 36); total number of words recalled following a 30-minute delay (max 12); and a discrimination index score representing a participant's ability to discriminate between old and new list items (max 12) and the Paired Associate Learning (PAL) subtest from the Cambridge automated neuropsychological test assessment battery (CANTAB; (Swainson et al. 2001). The score of interest for the latter measure was the number of pattern-position errors made at the six-pattern level.

Healthy and depressed control participants, as well as aMCI and a subgroup of early AD patients also completed the Rey Complex Figure Test (RCFT; copy, immediate and delayed recall trials; (Rey 1941). The measure of interest for this analysis was the 30-minute delayed recall condition.

4.4.3.2.3 Attention and executive function

All participant groups completed the trail making tests, Parts A and B (TMT A and B; (Reitan 1985). TMT A serves as a measure of psychomotor processing speed, while TMT B also adds a divided attention component. By subtracting the time to completion for TMT A of this test from TMT B, a measure of the 'executive' functioning component can be acquired independently of processing speed.

4.4.3.3 Statistical Analyses

Data was analysed using SPSS 14.0 for Windows. Z-scores were calculated for lexical and semantic fluency measures. A FD score was calculated for each participant by subtracting lexical fluency (P words) z-score from semantic fluency (animal words) z-score. Demographic variables and cognitive performance were examined using one-way group wise ANOVAs. As the group sizes were not equal, Tukey pair wise comparisons were carried out on all significant analyses when the assumption of homogeneity of variance was met, and Games–Howell pair wise comparisons were carried out when this assumption was violated. Reaction time data (e.g. TMT A and B) was log transformed prior to analysis in order to increase normality.

4.4.4 Results

4.4.4.1 Participant characteristics

Demographic characteristics are presented in Table 4.7. Groups did not differ in terms of age [F (3, 120) = 1.7; p =.17] or predicted pre-morbid IQ [F (3, 105) = 2.2; p = .10]. The estimated pre-morbid level of general intellectual functioning fell within a high average range for all four groups. There was a significant gender imbalance (F > M) in the depressive symptom group only (x2 = 5.6; p = .02; in all other cases p > .05). Patients with early stage AD performed at a lower level than all other groups on cognitive screening measures (MMSE: [F (3, 120) = 30.6; p < .0001]; ACE: [F (3, 120) = 30.6; p < .0001]); in all cases p < .0001. Total ACE scores of the aMCI group fell between that of the controls and early AD (MCI vs. C: p < .0001; MCI vs. AD: p < .0001) and was significantly different from both of these groups. By contrast, mean MMSE scores for control and aMCI groups did not differ. Mean total scores on both cognitive screening measures (MMSE and ACE) failed to discriminate the depressive control group from aMCI or healthy control participants.

Table 4.7 Means (SD) of clinical and demographic data

Variable	Controls (C; N = 24)	Depressed (D; N = 18)	aMCI (MCI; N = 47)	Early AD (AD; N = 35)	Statistic	Post hoc group differences (Tukey; Games-Howell) p < .05
Age	70.8 (7.8)	73.3 (6.3)	73.9 (6.4)	74.4 (6.6)	F(3, 120) = 1.7; p = .17	
Gender	9 M: 15 F	4 M: 14 F*	18 M: 29 F	17 M: 18 F	Chi-square = 3.5; p = .32	
NART FSIQ	118.5 (3.3)	116.2 (5.6)	116.8 (7.7)	113.5 (8.9)	F(3, 105) = 2.1; p = .10	
MMSE	28.9 (1.1)	28.5 (1.5)	28.3 (1.5)	24.9 (2.8)	F(3, 120) = 30.6; p < .0001	C,D,MCI > AD
ACE	94.5 (3.2)	91.3 (5.7)	89.4 (5.6)	75.5 (7.4)	F(3, 120) = 64.5; p < .0001	C > MCI; C, D, MCI > AD

AD, Alzheimer's disease; aMCI, amnestic mild cognitive impairment; M, male; F, female; NART FSIQ, National Adult Reading Test Full Scale IQ; MMSE, mini mental state exam; ACE, Addenbrookes cognitive examination

In keeping with proposed aMCI criteria (Petersen et al. 1999) patients with aMCI performed more than one standard deviation below age means on at least two formal measures of episodic verbal and visual memory, such as the HVLT total and delayed recall, the PAL task from the CANTAB, and the RCFT delayed recall.

4.4.4.2 Fluency performances according to patient group

4.4.4.2.1 Semantic fluency

Mean scores and standard deviations for semantic fluency, lexical fluency, and semantic/lexical discrepancy are presented for each of the four patient groups in Table 4.8.

As predicted, the mean semantic fluency score for the early AD group was significantly lower than all other patient groups (in all cases $p < .0001$). The aMCI patient group also generated significantly fewer animal names than did healthy controls ($p < .0001$). The mean number of animals generated by both the aMCI and depressed groups fell between that of early AD and control participants and was significantly different from both of these groups (MCI vs. C: $p < .0001$; MCI vs. AD: $p < .0001$; D vs. C: $p = .04$; D vs. AD: $p < .0001$).

4.4.4.2.2 Lexical fluency

Lexical fluency scores were higher for the aMCI and healthy control groups than for the early AD patients (C vs. AD: $p = .02$; MCI vs. AD: $p = .002$). No other group differences in lexical fluency performance were observed.

4.4.4.2.3 Fluency discrepancy scores

The mean FD scores for aMCI and early AD patients were significantly higher than those of healthy control (MCI vs. C: $p < .0001$; AD vs. C: $p < .0001$) and depressed control groups (MCI vs. D: $p = .01$; AD vs. D: $p = .001$), but did not differ significantly from each other (MCI vs. AD: $p = .72$) (see Table 4.8). The negative mean discrepancy scores of the aMCI and early AD groups indicate that, on average, these patients generated fewer animals than 'P' words within the 1-minute time frame. The opposite pattern (animals > P words) was observed for the healthy controls and the depressive control group, indicated by a positive mean score. Mean FD scores did not differ between the two control groups (C vs. D: $p = .27$). To assess the usefulness of FD scores in classifying patients in comparison to a more commonly used word list learning task, receiver operating characteristic (ROC) curves were constructed (Figures 4.10 – 4.12). Figures 4.10 and 4.11 demonstrate that the FD index has

greater sensitivity and specificity in distinguishing healthy and depressed control participants from those with aMCI, whereas delayed verbal recall ability retains superiority in the early AD versus healthy control comparison (Figure 4.12). The area under the curve for each of the three comparisons is as follows: C versus MCI: 0.84 (FD) and 0.78 (HVLT-delay); D versus MCI: 0.75 (FD) and 0.71 (HVLT-delay); C versus AD: 0.85 (FD) and 0.94 (HVLT-delay).

Table 4.8 Means (and SDs) for verbal fluency and other relevant measures of episodic memory, processing speed and executive function

Variable		Controls (N = 24)	Depressed (N = 18)	aMCI (N = 47)	Early AD (N = 35)*	Statistic	Post hoc groups differences (Tukey; Games-Howell) p < .05
Animal fluency	Raw score Z score	21.3 (5.8) 0.00 (1.00)	17.6 (4.8) -0.65 (0.84)	15.4 (4.3) -1.03(0.75)	10.4 (3.3) -1.89 (0.57)	F(3, 120) = 30.1; p < .0001	C > MCI,D> AD
P word fluency	Raw score Z score	15.7 (5.7) 0.00 (1.00)	14.3 (5.4) -0.23 (0.95)	15.9 (4.3) 0.04 (0.75)	11.9 (4.4) -0.66 (0.76)	F(3, 120) = 5.2; p <=.002	C, MCI > AD
Fluency discrepancy	z score	0.00 (0.86)	-0.42 (0.69)	-1.07(0.67)	-1.24 (0.79)	F(3, 120) = 17.1; p < .0001	C, D > MCI, AD
HVLT-R delay	Raw score	8.1 (2.7)	7.4 (3.6)	4.8 (3.3)	1.4 (2.8)	F(3, 103) = 19.8; p < .0001	C, D > MCI, AD
RCFT delay	Raw score	16.9 (6.8)	15.3 (6.4)	11.4 (6.9)	2.4 (3.9)	F(3, 96) = 16.3; p <.0001	C, D, MCI > AD; C > MCI
PAL 6 box errors	Raw score	7.9 (6.7)	11.3 (7.6)	17.1 (14.5)	40.8 (11.1)	F(3, 109) = 38.9; p < .0001	AD > C, D,MCI; MCI > C
TMT A+ (s)	Raw score	40.2 (10.5)	53.3 (22.6)	45.1 (18.9)	59.0 (22.2)	F(3, 104) = 4.2; p = .007	AD > C, MCI
TMT B+ (s)	Raw score	88.7 (30.7)	140.2 (53.7)	117.1(66.1)	183.3 (96.4)	F(3, 106) = 8.9; p < .0001	AD > C, MCI; D > C
TMT B- TMT A+ (s)	Raw score	48.5 (24.6)	86.9 (49.1)	72.0 (60.4)	92.0 (56.8)	F(3, 99) = 2.8; p = .046	

AD, Alzheimer's disease; aMCI, amnestic mild cognitive impairment; HVLT-R delay, Hopkins verbal learning test-revised delayed recall; RCFT delay, Rey complex figure test delayed recall; PAL 6 box error, total number of errors made at the 6 pattern stage of the paired associated learning subtest from the CANTAB visual memory battery; TMT, trail making test; *AD group sample size varies by analysis; + Log transformed in order to increase normality of data set

4.4.4.3 Episodic memory

By definition, the aMCI group performed at significantly lower levels than healthy controls on a number of verbal and visual episodic memory measures: HVLT total and delayed recall scores (total score [means (SD) for controls/patients]: 23.4 (5.1)/19.3 (4.6); t(2; 67) = 3.4, p = .01 and delayed recall: 8.1 (2.7)/4.8 (3.3); t(2, 67) = 4.2, p < :001); the PAL task from the CANTAB (7.9 (6.7)/17.1 (14.5); t(2, 68) = -3.7, p < .001); and the RCF delayed recall (16.8 (6.8)/11.4 (6.9); t(2, 66) = 3.1, p = .003).

Figure 4.9 Means and distribution of the differences (animals minus P-word) of verbal fluency z scores in diagnostic groups.

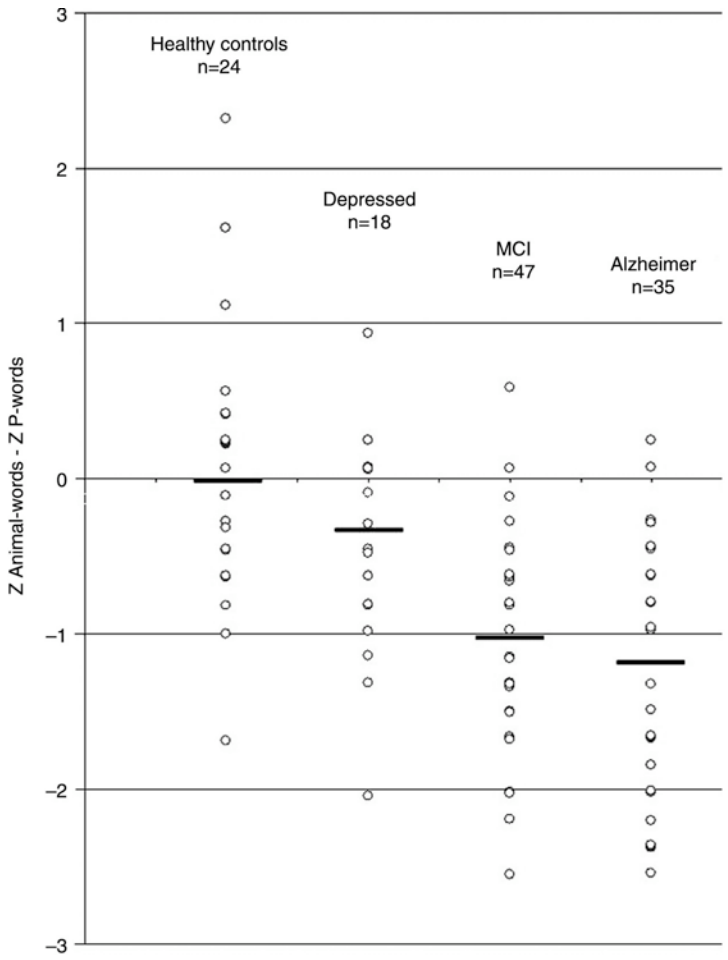
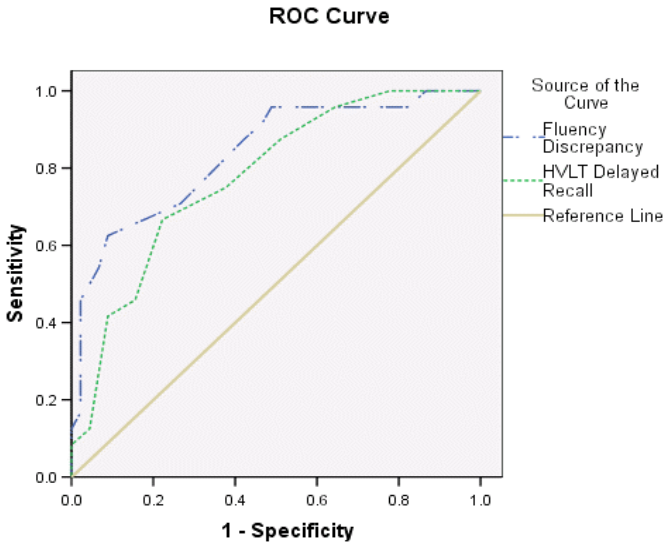
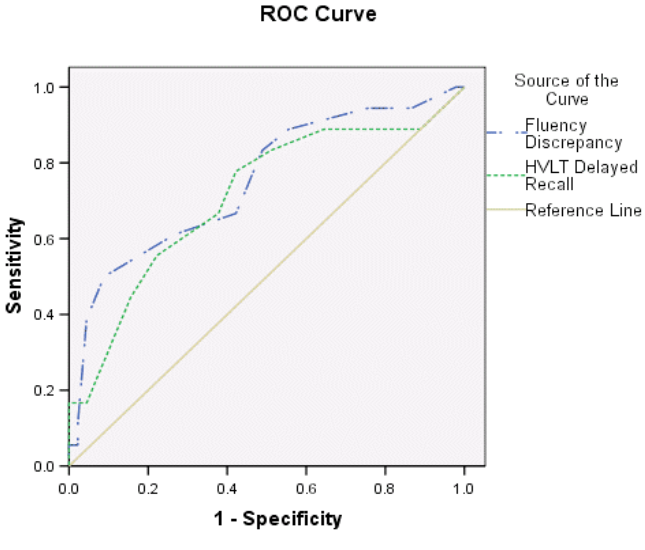


Figure 4.10 Receiver Operating Characteristic (ROC) curve for differentiating aMCI participants from healthy controls.



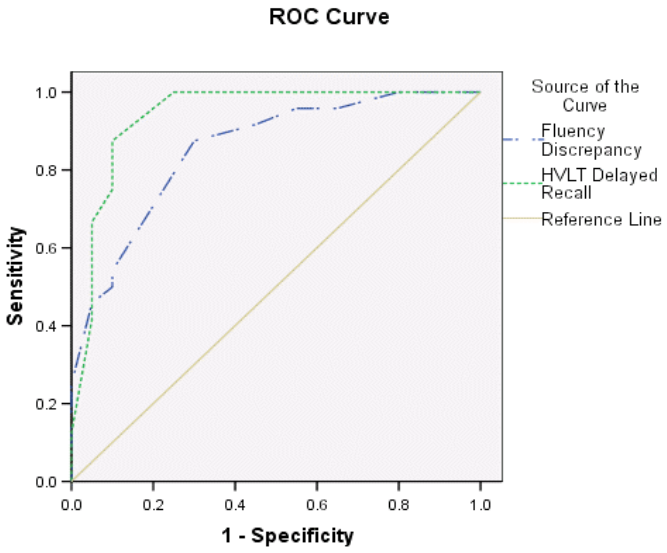
Fluency Discrepancy: AUC = 0.837, p<0.0001; HVLT Delayed Recall: AUC = 0.777, p <0.0001

Figure 4.11 Receiver Operating Characteristic (ROC) curve for differentiating aMCI participants from depressed controls



Fluency Discrepancy: AUC = 0.751, p=0.002; HVLT Delayed Recall: AUC = 0.710, p =0.009

Figure 4.12 Receiver Operating Characteristic (ROC) curve for differentiating Alzheimer's disease participants from healthy controls.



Fluency Discrepancy: AUC = 0.854. p<0.0001; HVLT Delayed Recall: AUC = 0.940, p <0.0001

4.4.4.4 Attention and executive function

The early AD group were slower than healthy controls and aMCI patients to complete the TMT A (C vs. AD: p = .01; MCI vs. AD: p = .03). Furthermore, early AD and depressive control patients were on average slower than healthy controls to complete Part B of the TMT (C vs. AD: p < .0001; C vs. D: p = .02). A TMT B–A measure was calculated by subtracting the time to completion for Part A of the TMT from Part B of the same task. On analysis, there proved to be a marginally significant group difference (F (3, 99) = 2.8; p = .046); however none of the multiple comparison procedures reached significance.

4.4.5 Discussion

Our findings demonstrate that the pattern of performance on lexical and semantic fluency tasks is distinctly different in early AD and aMCI compared with healthy and depressive age-matched controls. Specifically, early AD and aMCI patients show a greater magnitude of

impairment in semantic, as compared to lexical fluency, relative to healthy age-matched controls (Lezak 2004;Spreen and Strauss 1998).

The findings are consistent with longitudinal data demonstrating that semantic fluency deficits pre-date the lexical fluency impairment seen in the preclinical stages of AD (Auriacombe et al. 2006) and with cross-sectional data showing an impairment of semantic fluency in the absence of any lexical fluency deficit in aMCI (Alladi et al. 2006;Dudas et al. 2005b;Lonie et al. 2008;Murphy et al. 2006). In cases where a lexical fluency deficit has been documented in aMCI or preclinical AD, the general level of cognitive function of the aMCI patients is often comparatively low, for example a mean MMSE of 24.5, compared to our early AD patient group of 24.9 (Jones et al. 2006).

Our results implicate early failure of one or more aspect(s) of the semantic memory system in aMCI and early AD. They do not directly address the question of which mechanisms of the semantic memory system may be at fault. The relatively sound performance of our aMCI patients on additional measures of verbal initiation and speeded divided attention would suggest that these specific aspects of executive function do not underlie the semantic fluency deficit. Similarly, Auriacombe and colleagues (Auriacombe et al. 2006) failed to implicate executive processes in the verbal fluency decline characterising AD. It remains possible that the semantic fluency deficit seen in aMCI and early AD reflects degradation of amodal representations forming the semantic memory store, or failure of the executive processes that help to direct and control semantic activation of this store (Jefferies and Lambon Ralph 2006). Detailed quantitative and qualitative longitudinal investigation of semantic memory function in an aMCI cohort will be required to differentiate between these possibilities.

What appears increasingly clear, nonetheless, is the co-existence of cognitive deficits in domains other than episodic memory function in aMCI; most notably semantic memory deficits (Adlam et al. 2006;Alladi et al. 2006;Lam et al. 2006;Murphy et al. 2006;Perry and Hodges 2000;Riberio et al. 2006;Vogel et al. 2005). The original MCI criteria have been modified to some extent to reflect this (Petersen, 2004). FD scores performed as well as, if not better than delayed verbal recall measures in distinguishing between patients with aMCI or early AD and age-matched controls or depressive controls. The discrepancy score may

therefore be especially sensitive to semantic memory failure in aMCI. There are a number of reasons why this might be so.

Verbal fluency scores are influenced by age and IQ (Lezak 2004;Spreen and Strauss 1998;Tombaugh et al. 1999), and performance on the two types of fluency measures is known to correlate (Tombaugh et al. 1999). Applying generalised cut-off scores at an individual patient level could conceivably reduce the sensitivity of fluency measures to deficits of a small magnitude. Lexical fluency performance, which remains comparatively well preserved in preclinical and very early AD (Alladi et al. 2006;Dudas et al. 2005b;Murphy et al. 2006) and is relatively insensitive to the effects of ageing in later life (Murphy et al. 2006), may act as a personalised benchmark, against which even subtle declines in an individual's semantic fluency performance can be measured. At least one study has shown an enhanced ability to detect cognitive deficits of a progressive nature using individualised IQ adjusted rather than group norms (Rentz et al. 2004). Based on a similar rationale is the use of regression equations to calculate expected performance levels for the individual patient (Crawford and Garthwaite 2006;Rentz et al. 2004).

An interesting observation was the comparative lesser ability of the FD score to discriminate early AD sufferers from healthy age-matched controls. Prior research with established AD sufferers suggests that the relative magnitude of phonemic and semantic fluency deficits remains constant across disease stages (Henry et al. 2004). This would imply that the discriminative power obtained for FD in the aMCI group should also be maintained in early AD. It is possible however that this argument does not hold in a preclinical disease phase; at which time the greater sensitivity of the discrepancy score may reflect semantic memory failure prior to disruption of wider executive processes.

There is continuing debate as to which of the two cognitive domains, that is, semantic memory abilities or executive processes, constitutes the secondary area of impairment in early and preclinical AD. It is conceivable that as the disease progresses, 'executive' aspects of semantic retrieval may add to poor performance due to existing semantic memory dysfunction on category fluency tasks. When this occurs, deficits on lexical tasks would also be expected to be present. As lexical fluency performance declines, the discrepancy in

fluency performance may become a less sensitive measure of cognitive dysfunction than category fluency performance or delayed recall alone.

In our study, the differential semantic fluency deficit was unique to early AD and aMCI patient groups suggesting that it may be of assistance in differentiating these groups from the healthy elderly or elderly patients with depressive symptoms. Other studies have found that comparison of performance on the two different types of fluency tests is helpful in distinguishing between certain dementia syndromes (Jones et al. 2006;Marczinski and Kertesz 2006). This level of discrimination may not be achievable by examination of fluency scores in isolation, or indeed by cognitive screening measures, both of which failed to discriminate between our aMCI and depressive control groups.

The use of a single letter and category (i.e. P and animals), as opposed to an average of three (i.e. FAS or animals, fruits, and vegetables), might be expected to yield a less reliable fluency score, and it will be necessary to replicate the above findings in other clinical samples using alternative categories and letters. One recent study, comparing the number of animals with F-words generated, has demonstrated a similar ability to discriminate early AD and aMCI patients from healthy controls. This study lacked the important inclusion of a depressive control group matched for general level of cognitive functioning and support needs (Murphy et al. 2006). Furthermore, the advantage of the measures used in the current study is that they are obtainable within the context of a brief cognitive screening measure (ACE and ACE-R), without necessity for the administration of any supplementary tests.

In order for FD scores to be of value in a differential diagnostic sense, they must facilitate discrimination between pathological and age appropriate performance at the individual patient level. Only one healthy control and one outpatient with depressive symptoms in our study obtained a higher lexical than category fluency score, but there was a good deal of directional variability in the early AD and aMCI patient groups. Hence the FD measure may prove to be of greatest assistance identifying those persons who are not likely to develop AD over time (negative predictive value). We are conducting longitudinal follow-up of the aMCI patients in order to determine whether or not this is the case.

Two previous studies have contrasted fluency difference scores in AD and healthy controls (Cerhan et al. 2002;Sherman and Massman 1999). In both, the overwhelming majority of AD patients demonstrated the expected semantic < lexical fluency performance pattern. However, the existence of a notable number of AD patients exhibiting the opposite pattern led both authors to conclude that the FD score may be of limited sensitivity.

In the present study, optimal cut-off scores can be generated from the receiver– operator curves in Figure 2. For example, a cut-off of 3.5 and higher for the animal minus letter-P scores (or a difference of z scores of 2 37%) gives a sensitivity of 63 and 50% to aMCI in healthy volunteers and participants with depressive symptoms, respectively, but a very high specificity of 91% (Table 3). On the other hand, a more lenient cut-off of greater or equal to 1.5 (z score of 2 72%) gives a sensitivity for aMCI or Alzheimer's dementia versus controls of 88%, with a reduced specificity of 58 and 70%, respectively.

4.4.6 Conclusion

Our results indicate that the differential semantic fluency deficit seen in AD is also present in aMCI. They provide support for the presence of semantic memory impairment in the preclinical and very early stages of AD and emphasize the need to broaden the current conceptualisation of aMCI beyond that of a purely amnesic state.

FD scores appear equally adept, if not superior, to episodic memory measures at identifying aMCI. Importantly, the presence of depressive symptoms in psychiatric outpatients does not appear to influence the relative performance on the two types of fluency measures, suggesting that consideration of FD may be of assistance in early differential diagnosis. Longitudinal follow-up of the aMCI will determine whether or not FD scores are of equal utility in a prognostic sense.

4.5 Dual Task Performance in early AD, aMCI and Depression.

(Published in part in Psychological Medicine. 2008; Vol 39; 23 - 31)

4.5.1 Abstract

Background

The dual task paradigm (Baddeley et al. 1986;Della Sala et al. 1995) has been proposed as a sensitive measure of Alzheimer's dementia, early in the disease process.

Aims

Our aim was to investigate this claim.

Methods

We administered the modified dual task paradigm (utilising a pencil-and-paper version of a tracking task) to 33 patients with amnestic mild cognitive impairment (aMCI) and 10 with very early Alzheimer's disease, as well as 21 healthy elderly subjects and 17 controls with depressive symptoms. All groups were closely matched for age and pre-morbid intellectual ability.

Results

There were no group differences in dual task performance, despite poor performance in episodic memory tests of the aMCI and early Alzheimer's disease groups. In contrast, the Alzheimer patients were specifically impaired in the trail-making test B, another commonly used test of divided attention.

Conclusions

The dual task paradigm lacks sensitivity for use in the early differential diagnosis of Alzheimer's disease.

4.5.2 Introduction

Alzheimer's disease (AD) is the most common form of dementia, estimated to rise dramatically in the future (Wimo et al. 2003). Research has focused on early accurate diagnosis and intervention. The construct amnestic Mild Cognitive Impairment' (aMCI; (Petersen et al. 2001) has become increasingly popular to predict those who are most at risk for developing dementia. It is considered a transitional stage between normal ageing and the earliest clinical diagnosis of AD (Petersen 2005c;Petersen and O'Brien 2006). Research on clinic-based samples has suggested that the conversion rate from aMCI to dementia is 10–15% per year (Petersen et al. 1999;Storandt et al. 2006) compared with between 1% and 2% in a normal age matched non-clinical sample.

While primary impairment in very early AD includes episodic memory function, many authors have reported that attention and executive functioning are also vulnerable at this stage (Parasuraman and Haxby 1993;Perry and Hodges 1999). In particular, people with early AD exhibit marked difficulty dividing their attention between two concurrent tasks. By comparing performance of a synchronous dual task with that of identical task components done separately and consecutively, a deficit in dual performance can be attributed to failure of the central executive that coordinates the simultaneous operation of these components (Baddeley et al. 1986). One advantage of the dual task paradigm is that it avoids modality-specific interference between tasks: the tracking task is presented visually and a manual response is required; information for the digit span task is presented aurally with a verbal response (Nebes et al. 2001) A further strength is that task demands can be fixed at individual ability levels, controlling for individual variation in performance in the component parts of the dual task. Therefore, each patient is his or her own control, adjusting for the generally poorer performance of AD patients in the baseline tasks (Logie et al. 2004).

Research has suggested that failure of the 'coordination' function is characteristic of mild AD in a laboratory setting. Participants with mild AD appear to be impaired, irrespective of task demands, and this impairment has been found to worsen with illness progression (Baddeley et al. 1986; (MacPherson et al. 2004). Proponents of the dual task paradigm suggest such findings are in contrast to normal ageing, which they believe has a relatively minor effect on dual task performance (e.g. (Baddeley et al. 1986;Hartley and Little

1999;Logie et al. 2004); but see (Crossley and Hiscock 1992). The equipment used for this test is often an expensive computerized tracking device impractical for clinical settings (e.g. (Baddeley et al. 1991;Logie et al. 2004). Della Sala et al. (Della Sala et al. 1995) developed a modified pencil-and-paper version of the tracking component for the dual task. This has been reported to produce results comparable with the original instrument (Della Sala et al. 1995;Sebastian et al. 2006). To our knowledge the dual task paradigm has not been investigated with a sample defined according to recent aMCI criteria (Petersen et al. 1999).

The aim of this study was therefore to assess dual task performance in aMCI to ascertain whether this measure can be useful in the early diagnosis of AD. As AD is associated with a specific impairment in the aspect of working memory that coordinates performance of two separate tasks, we predicted that the performance of people with aMCI and very early AD should be significantly lower than that of aged matched controls. Furthermore, the inclusion of a group of elderly patients with symptoms of depression would test the specificity of dual task impairments in AD. On the basis of the previous research, we predicted that the depressed group would show impairment in the dual task compared with controls.

4.5.3 Method

4.5.3.1 Participants

We examined 33 patients with aMCI, 10 early AD patients, 17 control out patients with depressive symptoms and 21 healthy elderly controls. Thirteen aMCI participants showed an impairment of more than 2 standard deviations (SD) below our control mean on two or more tests, a further four showed impairments of 1.5 to 2 SD on two or more tests, 12 participants were impaired at 1–1.5 SD below control means on two or more tests, and the final four participants performed more than 1 SD below controls on one episodic memory test. Twenty four of the 33 aMCI underwent some form of neuroimaging before or during the study period in order to exclude non-neurodegenerative causes for their cognitive impairment. Mini Mental State Examination (MMSE) scores for the aMCI group ranged from 24 to 30, with a mean of 28.3. The final aMCI group consisted of 15 males and 18 females with a mean age of 73.3 years (range 58–85 years).

For the healthy elderly control group (MMSE 28–30), we recruited spouses or carers of patients who had attended the service. The healthy elderly control group was matched as closely as possible to the aMCI and early AD groups in terms of age and estimated pre-morbid intelligence quotient (IQ). The final elderly control group consisted of eight males and 13 females with a mean age of 69.5 years (range 59–81 years).

Ten participants diagnosed with AD, in accordance with National Institute of Neurologic, Communicative Disorders and Stroke–AD and Related Disorders Association (NINCDS-ADRDA; (McKhann et al. 1984) and DSM-IV diagnostic criteria, took part in the current study. All early AD patients scored above 23/30 on the MMSE and above 65/100 on the more comprehensive Addenbrookes Cognitive Examination (ACE; (Mathuranath et al. 2000)), indicating relatively mild disease. The final early AD group consisted of three males and seven females with a mean age of 73.6 years (range 65–81 years).

Seventeen participants with depressive symptoms (MMSE 25–30) were recruited via local psychiatric out-patient clinics and day hospitals. In an attempt to match this patient group with the aMCI group in terms of illness severity, patients with milder forms of depression were included. Fifteen of the 17 participants were receiving treatment for their symptoms at the time of testing; all but two of these pharmaceutical in nature. As it has been suggested that type of depression does not influence the magnitude of cognitive deficits (Christensen et al. 1997), participants with a variety of disorders were included. Eight patients had a history of major depression, two of bipolar disorder, two were suffering from anxiety disorders with depressive features, three were considered Dysthymic and two were considered to be suffering with a subclinical level of depressive symptoms. Mean geriatric depression scale (30-item version) score for this group was 13.2 (range 0–27). The group consisted of three males and 14 females with a mean age of 73.3 years (range 65–84 years).

4.5.3.2 Neuropsychological tests

All participants completed a variation on the modified dual task paradigm (Della Sala, personal communication, 2008; (Della Sala 2005)) together with the MMSE, ACE, NART, HVLT-R, PAL, TMTA & B. Further details of the method of subject recruitment, participant characteristics together with psychometric and administrative characteristics of

the cognitive screening measures employed can be found within Chapter 2, Materials and Methods, sections 2.1 and 2.

The primary measure of interest was the combined dual task score as denoted by μ

$\mu=(1-[(Pm+Pt)/2])*100$

where m is the combined dual task score, Pm is the proportional loss in span performance between single (Xsingle) and dual task (Xdual) conditions, [(Xsingle – Xdual)/Xsingle] while Pt is the equivalent proportional loss in tracking score. Thus a score of 100 would represent no dual task decrement and lower scores reflect greater dual task decrements.

Each of these measures has been shown to be sensitive to very early AD (Blackwell et al. 2004;Chen et al. 2000;Hogervorst et al. 2002;Nathan et al. 2001;Stockholm et al. 2006;Swainson et al. 2001). Neuropsychological assessments lasted approximately 90 min in total. The order of test administration was identical for all assessments.

4.5.3.3 Statistical Analysis

Data were analysed using SPSS 12.0 for Windows (SPSS Inc., Chicago, IL, USA). Demographic variables were analysed using univariate analysis of variance (ANOVA), and Tukey honestly significantly different pair wise comparisons were carried out on all significant analyses where possible. Where the assumption of homogeneity of variance was not met, this was adjusted for using Games–Howell post-hoc pair wise comparisons, given that the sample sizes were unequal in the current analysis. A univariate ANOVA was carried out on the overall decrement score (see above). Decrement scores broken down into tracking decrement and digit span decrement were also calculated and examined using ANOVAs. Two participants in the early AD group were incapable of completing the TMT B; in these cases a default ceiling score of 500 seconds to completion was applied.

4.5.4 Results

Table 4.9 Means (and SDs) for demographic data and cognitive screening measures

Variable	Controls (n=21)	aMCI (n=38)	Depression (n=17)	Early AD (n=10)	Group Differences
Age	69.5 (7.3)	73.1 (6.3)	73.3 (6.6)	73.60 (5.8)	ns
Gender	8M:13F	16M:22F	3M:14F	3M:7F	ns
NART	118.2 (2.9)	116.5 (8.1)	116.8 (6.2)	115.60 (5.5)	ns
MMSE+	29.10 (.7)	28.47 (1.6)	28.59 (1.5)	25.00 (2.3)	CT=MCI=Dep>AD
ACE	94.57 (3.3)	89.87 (5.9)	91.71 (5.0)	76.70 (6.6)	CT > aMCI>AD Dep>AD

+ Games-Howell Multiple Comparison carried out due to lack of HOV; M = Male; F = Female; AD = Alzheimer's Disease; aMCI = Amnesic Mild Cognitive Impairment; ACE = Addenbrookes Cognitive Examination.

4.5.4.1 Participant characteristics

Demographic matching characteristics are presented in Table 4.9. There were no group differences in age [$F(3, 77)=1.73$] or estimated pre-morbid full-scale IQ [$F(3, 75)=0.55$]. The mean MMSE score for the early AD group was, as expected, significantly lower than that of the other groups [$F(3, 77)=17.70$, $p<0.0001$] (AD v. healthy controls, $p=0.001$; AD v. controls with depressive symptoms, $p<0.005$; AD v. controls, $p<0.005$). No other group differences in mean MMSE score were noted. As expected, the early AD patients had significantly lower mean ACE scores than did all other groups [$F(3, 77)=29.30$, $p<0.0001$] (post-hoc tests as above in all cases, $p<0.0001$). The ACE also discriminated between normal elderly control participants and aMCI patients, with the latter group obtaining a significantly lower mean ACE score (post hoc $p=0.01$).

4.5.4.2 Dual task performance

Group means and SDs for the digit span task and the tracking measures of the modified dual task paradigm are presented in Table 4.10. Mean percentage scores for performance on the concurrent tasks, the digit span tasks and the visuospatial tracking tasks for each of the four groups are presented in Table 4.11. On carrying out a one-way non-repeated ANOVA on the

overall decrement score, no group difference was found [F(3, 77)=0.63]. Similarly, no significant group differences were found for any of the other component tasks or decrement scores.

Table 4.10 Digit span and individual component measures of the dual task (span and tracking, performed separately and together)

Task	Controls (n = 21)	Depression (n = 17)	aMCI (n = 33)	Early AD (n = 10)
Digit span	5.5 (0.7)	5.8 (1.0)	5.6 (0.9)	5.1 (0.7)
Digit span (single) a	1.0 (0.03)	0.9 (0.05)	1.0 (0.05)	1.0 (0.03)
Digit span (dual) a	0.9 (0.05)	0.9 (0.08)	0.9 (0.08)	1.0 (0.02)
Tracking (single) b	141 (56.5)	140 (51.7)	126 (38.9)	120 (46.3)
Tracking (dual) b	122 (46.0)	126 (58.3)	114 (36.3)	107 (35.6)

aMCI, Amnestic mild cognitive impairment; AD, Alzheimer's Disease; Values are given as mean (standard deviation); a = Proportion of digits recalled in the correct position (1 = all correct); b = Number of circles joined in 90 second.

Table 4.11 Percentage loss of performance in component tasks and overall decrement score during the dual task

Task	Controls (n = 21)	Depression (n = 17)	aMCI (n = 33)	Early AD (n = 10)
Digit span	96 (3.8)	97 (8.6)	98 (7.6)	100 (3.3)
Tracking	90 (22.8)	88 (16.8)	92 (15.4)	93 (17.6)
Overall decrement	93 (11.1)	92 (8.2)	95 (8.6)	97 (9.1)

aMCI, Amnestic mild cognitive impairment; AD, Alzheimer's disease; Values are given as mean (standard deviation); a percentage loss of performance scores were calculated as (1x[(XsinglexXdual)/Xsingle])*100 and the overall decrement score as mu=(1-[(Pm+Pt)/2])*100, as described in the Method section.

4.5.4.3 Other cognitive functions

Group mean scores and SDs for the HVLT-R, the number of errors at the six pattern level of the PAL and the TMT B are presented in Table 4.12. On analysing the HVLT-R delayed recall data, there was a significant group effect [F(3, 77)=12.39, p<0.0001]. On closer analysis, the AD group recalled significantly fewer words than the healthy control

($p<0.0001$) and depression groups ($p<0.0001$). Similarly, the aMCI group performed more poorly than the healthy control ($p<0.005$) and depression groups ($p<0.01$). No significant difference was found between the AD and aMCI groups. The performance of the elderly control and depression groups on the HVLT-R delayed recall did not differ. However, the AD group made significantly more errors at the six pattern stage of the PAL compared with all other groups [$F(3, 755)=22.82$, $p<0.0001$] (post hoc tests comparing AD with other groups were in all cases $p<0.0001$). The aMCI group's error scores fell between those of the healthy control and AD groups, and significantly differed from both of these (aMCI v. healthy controls, $p<0.05$; aMCI v. AD, $p<0.0001$). A significant group effect was also found for the TMT B [$F(3, 77)=8.62$, $p<0.0001$]. In the post hoc analyses, only the control and depression groups differed in terms of TMT B scores ($p<0.05$); participants with depressive symptoms took significantly longer to complete the task. However, once time to completion on TMT part A (a measure of psychomotor speed) was statistically controlled for, a different pattern of group differences emerged [$F(3, 76)=7.76$, $p<0.0001$]. Specifically, it was found that the participants with AD took longer to complete the second task compared with all other groups (aMCI v. AD, $p<0.0001$; healthy controls v. AD, $p<0.0001$; controls with depressive symptoms v. AD, $p<0.05$). The group difference between the control and depressive symptom groups was no longer significant. No other group differences were uncovered.

Table 4.12 Means (and SDs) for supplementary neuropsychological measures

Task	Controls (n = 21)	Depression (n = 17)	aMCI (n = 33)	Early AD (n = 10)	Group differences
HVLT-R delay	8.1 (2.8)	8.1 (3.3)	4.9 (3.3)	2.1 (3.7)	Controls = depression > aMCI = AD
PAL errors a	7.8 (6.9)	10.9 (7.8)	16.5 (12.9)	40.7 (10.6)	Controls, depression, aMCI < AD Controls < aMCI
TMT A	40.3 (11.2)	54.1 (23.1)	49.6 (36.1)	57.6 (25.3)	-
TMT B	87.6 (31.5)	134.2 (53.6)	106.3 (49.4)	216.7 (157.7)	Controls < depression Controls, depression, aMCI < AD b

a Games-Howell Multiple Comparison carried out due to lack of homogeneity of variance; b After removing effects of TMT A (see text); AD = Alzheimer's Disease; aMCI = Amnesic Mild Cognitive Impairment; HVLT-R = Hopkins Verbal Learning Test-Revised; PAL Errors = 6 Pattern Stage errors from the Paired Associates Learning test; TMTA = Trail Making Test Part A; TMT B = Trail Making Test Part B.

4.5.5 Discussion

This study investigated the claim that the dual task paradigm can be used in the early diagnosis of dementia of the Alzheimer's type. We assessed the concurrent performance of a visuospatial tracing task and a digit span forward task in four diagnostic groups with aMCI (MMSE 24–30), early AD (MMSE 23–29), symptoms of depression (MMSE 25-30) and healthy elderly controls (MMSE 28–30). Our results show that aMCI is not associated with impaired dual task performance; those with aMCI had comparable performance to healthy older adults and older adults with depressive symptoms. Our early AD group was similarly unimpaired on the modified dual task paradigm relative to depressive and non-depressive elderly control groups and the presence of depressive symptoms appeared to have no effect on dual task performance. By contrast, and indeed by definition, episodic memory impairments were present in the aMCI and early AD groups. The early AD group also exhibited an impaired ability to divide their attention at pace, as indicated by part B of the TMT.

These results shed some light on previous findings. One line of research has suggested that dual task performance is vulnerable to the influence of AD, even early in the disease course (Baddeley et al. 2001;Logie et al. 2004). However, such studies generally involve participants varying in severity from minimal to mild AD. When participants with AD are divided by severity using the MMSE, only the more severely ill patients (e.g. MMSE <24) are impaired on the dual task paradigm (Crossley et al. 2004;Greene et al. 1995;Perry et al. 2000). This result is in agreement with the absence of impairment on the dual task measure observed in the current study in early AD. The combined findings suggest that dual task impairments are generally not observed early on in the AD process, with MMSE scores above 23/30.

Only one other study has investigated the dual task performance of a group of older adults with cognitive impairment without a diagnosis of dementia (Holtzer et al. 2004). Cognitively impaired adults, defined by a dementia rating scale (DRS) cut-off score of <124 (Mattis 1988), performed two tasks in different modalities at the same time. Two combinations of tests were used: a visual cancellation task (where participants were required to cross out a specified stimulus type from a field of stimuli) combined with a digit span task, and the same

visual cancellation task combined with a verbal fluency task. The researchers report that their cognitively impaired group exhibited a significantly larger dual task decrement than age matched controls. However, the cognitively impaired group in the Holtzer et al. (Holtzer et al. 2004) study was identified solely on the basis of a DRS cut-off score falling at or below levels that are indicative of an underlying dementia. It is for this reason difficult to be certain of, or to compare, disease severity of this 'minimally cognitively impaired' group with other studies, which commonly use well-established clinical and research criteria to define patient groups. Furthermore, the cognitively impaired group in the Holtzer et al. (Holtzer et al. 2004) study, were significantly less well educated than the control groups, while in the current study, participant groups were well matched both in terms of age and estimated levels of pre-morbid intelligence.

Holtzer et al. (Holtzer et al. 2004) did not investigate the potential influence of depression on dual task performance. This is crucial where consideration is being given to the early and differential diagnostic value of a neuropsychological measure. Hasher & Zacks (Hasher and Zacks 1979;Mattis 1988) confirmed our result that people with depression show impaired attention during effortful processing tasks, for instance on measures of divided attention such as the TMT B (Mahurin et al. 2006;Nathan et al. 2001). Only one study has investigated the effect of depressive symptoms on Baddeley et al.'s (Baddeley et al. 1986) original dual task paradigm (Nebes et al. 2001). This indicated that people with depression had a significantly greater decrement in computerised tracking performance and a composite decrement measure than non-depressed controls. No study to date has investigated the effects of clinically depressed mood on the modified version of the dual task paradigm to replicate or contradict our negative result (Della Sala et al. 1995). A strength of the current investigation relates to the availability of additional neuropsychological data demonstrating the existence of significant episodic memory impairments in aMCI and early AD and additional impairment of speeded divided attention (as assessed by TMT B) in early AD. The TMT B assesses the ability to divide attention back and forth between multiple lines of thought (connecting numbers and letters, respectively), but differs from the dual task paradigm in that its different components are not drawn from separate modalities. Performance is thus more vulnerable to reduced processing capacity. Several previous studies have demonstrated that TMT B is impaired in the very early and even pre-clinical stages of AD (Alladi et al. 2006;Baddeley et al. 1986;Baudic et al. 2006;Crossley et al. 2004;Crowell et al. 2002;Lafleche and Albert 1995;Nathan et al. 2001;Perry et al. 2000), although its specificity for AD, as distinct from, for example, depression, has not been established.

The Holtzer et al. (Holtzer et al. 2004) study compared the dual task performance of minimally cognitively impaired participants only with their performance on tests comprising the single task conditions (i.e. visual cancellation, digit span and letter fluency). However, these tests are not, generally speaking, associated with impairments in very early and pre-clinical AD and it is therefore not surprising that they are insensitive to cognitive deficits in the minimally impaired group, as was the case in this study.

One important methodological feature may have influenced the current results: While those studies reporting general dual task impairment in early AD used both computerised and pencil-and-paper versions of the tracking task, only the modified version utilising the pencil-and-paper tracking task (Della Sala et al. 1995) has been used in studies that separated participants by symptom severity. Thus, while patients who are minimally affected do not show impairments on the modified version of the task, it remains possible that they would show impairments if the test were more taxing – for instance if the dual task paradigm included the original computerised version of the tracking task. This version of the task requires increased effort and attention, as participants are required to adjust to an external influence (i.e. the speed of the light dot on the screen) rather than working at a self-defined rate. It may therefore be sufficiently taxing to identify those who are not picked up by the more straightforward pencil-and-paper tracking task. However, the paper and- pencil version (as opposed to the computerised task) is more likely to be adopted for widespread use in clinical and research practice, which underscores the relevance of our negative result.

A further methodological issue is the variability of dual task administration, which can lead to difficulties comparing findings across studies. We administered each of the three trials in blocks of 90 seconds, while some previous studies set the trial time at 120 seconds (e.g. (Perry et al. 2000). Most dual task studies have utilised pencil-and-paper tracking tasks that required participants to cross out boxes on an A4-size sheet of paper to form a chain (e.g. (Baddeley et al. 1997). The current task required participants to trace a line through linked empty circles on an A3-size sheet. While the initial dual task paradigm involved recording the number of completely correct digit strings (Baddeley et al. 1986), many subsequent studies, including the current investigation, have calculated the number of digits recalled in the correct order for this measure. The significance of such alterations to dual task administration requires further investigation.

A partial alternative explanation for our negative result is that a majority of individuals forming our aMCI group may fail to convert to AD in the future. If this proves to be the case, then the absence of dual task impairment would not be surprising. The issue will be resolved through the longitudinal follow-up of participants with aMCI, currently underway. However, the sound performance of our early AD group on the dual task measure makes it more likely that the negative result for our aMCI patients is due to lack of test sensitivity rather than absence of underlying AD pathology. The impaired performance of the early AD group on an alternative popular measure of speeded divided attention implies that the dual task measure lacks sensitivity to very early changes of an attentional/executive nature in AD.

4.5.6 Conclusion

People with early AD and aMCI did not display impaired performance on the modified version of the dual task paradigm at a time when episodic memory, and in the case of early AD, speeded divided attention, were significantly impaired. The likely explanation is that the dual task paradigm is insufficiently sensitive for use as an adjunctive cognitive tool in the early diagnosis of AD. Future longitudinal research is needed to investigate the use of dual task tests of varying demand in aMCI and very early AD participants in an effort to determine the potential influence of task demands and complexity on performance.

5. Neuropsychological Test Performance Across Time

5.1 Test re-test reliability of neuropsychological measures in healthy elderly controls

5.1.1 Abstract

Background

In cases of very early AD and MCI, where associated functional declines are minimal and generalised intellectual decline is absent, it is often necessary to rely on psychometric evidence of cognitive decline in order to arrive at a diagnosis. It is for this reason, crucial, that neuropsychological measures that are employed to assess cognitive performance in MCI and early AD are not only sensitive enough to detect mild impairments but also reliable enough to facilitate detection of mild declines in cognition over time. Few previous studies have examined the stability of performance of the healthy elderly on neuropsychological measures across time.

Aims

We therefore sought to establish the re-test reliability of the study neuropsychological assessment battery in a group of healthy, age and IQ matched control participants, prior to analysing the longitudinal neuropsychological performance of the aMCI sample.

Methods

Sixteen healthy elderly control subjects re-sat the neuropsychological assessment battery on a second occasion, across an average 28 month interval. Baseline and re-test scores were compared by way of paired t-tests, where data were normally distributed, and the Wilcoxon test, where visual inspection together with Kolmogorov-Smirnoff Z tests revealed that data were not normally distributed and effect sizes were calculated where statistically significant differences in baseline and re-test scores were found. Test re-test reliability was assessed using Person's correlations for parametric data and Spearman's rho for non-parametric data.

Lower and upper 95% limits of agreement were calculated for the mean interval score for each variable, together with 95% upper and lower boundary confidence limits of difference values, in accordance with the methods for assessing agreement between measurement

proposed by Bland and Altman (Bland and Altman 1986). Repeatability coefficients were calculated where the baseline – re-test difference score did not reach significance.

Results

Significant improvements were apparent at retest on the ACE, HVLT-R total recall & RCFT delayed recall, whereas declines of a significant magnitude were noted in association with performance on the Category Fluency task and Part A of the TMT. The associated effect sizes for the above change scores were uniformly large, indicating the likely practical significance of these difference findings. The mean performances of the control group at baseline and follow-up did not differ significantly on the remaining neuropsychological measures.

Re-test reliabilities were for the most part medium to large. The magnitude of score change required to surpass the upper and lower 95% confidence limit ranged from a minimum of 1 additional word for the total recall measure of the HVLT-R to a maximum of 4 additional words for the BNT, and a loss of 1 word for the total recall measure of the HVLT-R and 6 points for the copy component of the RCFT, respectively.

Conclusion

In summary, reliable test performances were observed on a majority of neuropsychological tasks across a mean 28 month interval in our healthy elderly control group. Practice effects were present for several episodic memory measures together with a widely used cognitive screening measure (ACE) and these should be taken into account when interpreting change in performance on such measures over time. The ceiling effects (MMSE, ACE immediate recall, HVLT-DI, RCFT copy) and low re-test reliabilities (PAL, Dual Task) observed for several other tasks imply these measures may be less suited to monitoring cognitive performance and detecting small but significant changes in performance across time, in patients with mild cognitive difficulties. The significant overall decline in the performance of the control group on several timed tasks cautions against the use of processing speed measures for monitoring purposes in an elderly population, without reference to reliable change index scores.

5.1.2 Introduction

The concept of cognitive decline comprises a key component of MCI criteria (Petersen et al. 1999) and is central to the diagnosis of early AD. Reference points utilised to establish evidence of cognitive decline in clinical practice include estimates of pre-morbid levels of general intellectual functioning and prior performance psychometric measures. In cases of very early AD and MCI, where associated functional declines are minimal and generalised intellectual decline is absent, it is often necessary to rely on psychometric evidence of cognitive decline in order to arrive at a diagnosis. It is for this reason, crucial, that neuropsychological measures that are employed to assess cognitive performance in MCI and early AD are not only sensitive enough to detect mild impairments but also reliable enough to facilitate detection of mild declines in cognition over time.

We therefore sought to establish the re-test reliability of the study neuropsychological assessment battery in a group of healthy, age and IQ matched control participants, prior to analysing the longitudinal neuropsychological performance of the aMCI sample.

Few previous studies have examined the stability of performance of the healthy elderly on neuropsychological measures across time. Bird et al (Bird et al. 2004) reported high, one month test re-test reliabilities (r=0.92) for the GNT based on the performance of 106 normal adults with a mean age of 57 years and an estimated high average pre-morbid level of intellectual function. Reliable change indices corrected for practice effects of between -1.5 & +3.5 were provided, suggesting that the GNT may be a useful tool for monitoring even small cognitive changes. The authors (Bird and Cipolotti 2007) later reported excellent test re-test reliabilities (i.e. r= 0.92) and slightly larger reliable change indices (+ & - 4.5) for the GNT in a large neurological sample re-tested within a two year period.

The psychometric properties of the category fluency test, were also examined in the former study (Bird et al. 2004). Moderate test re-test reliabilities (r=0.56), together with a statistically significant practice effect in the order of 1.3 points and reliable change indices corrected for practice of -7.6 & +10.5 were reported, using the category 'animals'. Another study examining re-test reliability of the animal fluency task in British adult samples

reported standard error of prediction scores in the region of 4.33 words (Harrison et al. 2000). To my knowledge, no study has reported on the test re-test reliability of the category fluency task in older adults using the combined categories of 'animals', 'fruits' & 'vegetables'.

Three studies have examined the test re-test reliability of the HVLT. As the total number of words recalled across the three learning trials is common to both the original (HVLT) and revised (HVLT-R) versions of the test, data from both are considered here. Schrijnemaekers et al., (Schrijnemaekers et al. 2006) reported significant learning effects for the total number of words recalled across the three trials of version 1 of the HVLT across a 2-3 year period in healthy elderly controls but not for small groups of early AD or MCI study participants. Duff et al., (Duff et al. 2007) reported a mean increase of 1 word for the total number of words recalled across the three learning trials of the HVLT-R following a 3 month interval in a very small sample (n=8) of elderly MCI patients. Statistically significant 9 month stability coefficients (r = 0.50) have also been documented for the total number of words recalled across the learning trials of the HVLT, in a small group (n=45) of healthy elderly adults (Rasmusson et al. 1995).

Mitrushina and Satz (Mitrushina and Satz 1995) report an absence of practice effects on the 60 item version of the BNT, at one and two year intervals in a sample of 122 healthy 57-85 year old volunteers with a mean high average full scale IQ. Moderate test re-test reliabilities were documented across the initial and final one year intervals (r=0.62 & 0.65, respectively) and a higher re-test reliability was observed across the 2-year interval (r=0.89). Similarly, Zec et al., (Zec et al. 2005) reported small non-significant practice effects of between 0.21 and 0.19 words on the 60 item version of the BNT in a large sample of healthy community dwelling elderly with a mean age of 67 years who were re-tested at 9-15 month (n=353) and one and a half year intervals (n=540). Test re-test reliability between the first two BNT assessments undertaken by the sample was moderately high (r=0.76) and statistically significant for each of these intervals. Some minor variations from the standardised administration procedures for the BNT were noted, including omission of prompts and time limitations, for the latter study.

Flanagan et al., (Flanagan and Jackson 1997) re-assessed 31 healthy older adults with a mean age of 64 years across brief (i.e. 7-17 day) intervals on the BNT (60 item version). Although statistically significant improvements in performance at re-test were noted, scores were highly correlated (r=0.91) and the effect size for the mean difference score was relatively small (r=0.37), suggesting that the degree of improvement was of limited practical importance. In this study, 95% of participants scored within 2 points of their baseline BNT performance at re-test.

Thus whilst statistically significant practice effects have been reported within a small sample of healthy elderly on the BNT across relatively brief time intervals, the degree of improvement noted was of a small magnitude and questionable clinical significance and practice effects have not been reported across longer re-testing intervals amongst the healthy elderly.

Test re-test reliabilities as high as r=.890 and r=.83 were reported in the original standardisation sample for the MMSE (Folstein et al. 1975), in association with the same and different examiners, respectively. Near perfect 4 week test re-test reliabilities (r = .99) were reported for the sample of dementia patients in this study (Lezak 2004). Tombaugh (Tombaugh 2005) examined MMSE 3 month and 5 year re-test reliability and practice effects in 160 elderly persons (mean age 76 years) whose intact cognitive status was independently verified at follow-up. In this study, statistically significant practice effects were observed across the shorter (i.e. 3 months) but not longer (i.e. 5 year) intervals, where mean variation was of a magnitude of less than ½ a MMSE point and retest reliability was r = 0.65.

Duff et al., (Duff et al. 2007) reported little mean change (i.e. -0.4) in total FAS score in a very small group of 8 highly educated older adults with MCI (mean age 72 years) across a 3 month interval. Harrison et al., (Harrison et al. 2000) documented high test-retest reliability (r = .82) for the COWAT in 90 middle aged British adults across variable 1-8 week intervals. Only 60% of this sample displayed improvements in their performance at second assessment, suggesting that practice effects are by no means universal. Reliable change indices were not reported. A re-test reliability of r=.71 has also been reported for a sample of healthy elderly persons re-tested following a one year interval (Snow et al. 1988). The COWAT thus

appears to be a reliable measure across varying time intervals among the middle aged and healthy elderly.

Reported reliability coefficients for the TMT vary within the healthy elderly, with most above r = 0.6, several at r = 0.9 and more within the r = 0.8 range (Spreen and Strauss 1998). Within a small sample of highly educated elderly patients (mean age 72 years) with MCI. Duff et al., (Duff et al. 2007) reported a mean increase of 8.5 seconds to complete TMTA on a second testing occasion (following a 3 month interval) together with an average decrease of 20 seconds in completing part B at re-test. Practice effects are usually seen on both parts of the TMT with effects on Part A more likely to reach significance as a result of the greater variability in TMTB performances. An absence of practice effects has however been documented across longer i.e. one year re-test intervals in one study (Basso et al. 1999) and there is evidence to suggest that gains in performance as a result of previous exposure to Part B of the TMT may not be retained across longer (i.e. 3 month) time intervals (McCaffrey et al. 1992).

Acceptable to high levels of test re-test reliability (0.64 – 0.88), together with an absence of practice effects have been reported for the total number of errors made on the PAL subtest of the CANTAB battery, across one month re-test intervals (Fowler et al. 1995) in small samples of healthy elderly and early stage AD patients. This subtest has also been shown to be sensitive to small changes in episodic memory function across even relatively brief i.e. 6-12 month intervals (Fowler et al. 1997). Significant practice effects were, however, observed within a sample of Questionable Dementia sufferers, resulting in a mean reduction of 4 errors across all learning trials on the second testing occasion.

Three studies have reported on the test re-test reliability of the RCFT in older adult samples. Following a one year interval, Berry et al., (Berry et al. 1991) reported reliabilities of a moderate magnitude (.47-.59) for the immediate and 30 minute delay recall trials, and low reliability for the copy condition in a sample of 41 ‘normal’ elderly subjects. Mitrushina and Satz (Mitrushina and Satz 1991) similarly reported relatively a low test re-test reliability coefficient (.56-.68) for the copy version of the RCFT when administered to a group of elderly subjects thrice annually. Re-test coefficients for the 3-mintue delay condition ranged from 0.57 to 0.77 across the three annual probes.

It has been noted, that the limited ranges of scores for the copy condition (as function of ceiling effects in the healthy elderly), may serve to artificially reduce the magnitude of the test re-test correlation coefficients (Meyers and Meyers 1995). The authors report an absence of significant differences in scores on the copy component of the RCFT at 6 month re-test in a small (i.e. n-12) sample of 'normal' subjects and report re-test correlation coefficients of 0.76 for the immediate recall and 0.89 for the delayed recall components of the task. The potential for contamination of the incidental nature of the learning paradigm to affect re-test recall performance is also noted.

A relatively low test re-test reliability Pearson correlation coefficient of r=0.44 has been reported in association with the value of mu derived from the Dual Task performance of 33 persons (Baddeley et al. 1997). The reliability of the memory span component of the dual task was noted to be especially low (r=0.36) in comparison with the tracking component of the task (r=0.76). Demographic details of the re-test sample and the re-test interval are not, however provided.

There are no published test re-test reliability data for the EENT, or, to the best of my knowledge, for the ACE and GFT.

Table 5.1 Summary of test-retest reliability findings for neuropsychological measures comprising the research battery

	Primary author (year)	r	Re-test interval (months)	RCI	Sample (size)	Practice effect (s/ns)	Mean CS (ES)
GNT	*Bird (2004)*	0.92	1	(1.5-3.5)	HA (106)		
	Bird (2007)	0.92	24	(-4.5-4.5)	N		
Animal Fluency	*Bird (2004)*	0.56	1	(-7.6-10.5)	HA 106	1.3 (s)	
HVLT-R total recall	*Schrijnemaeker (2006)*		24–36		HE MCI AD	(s) (ns) (ns	
	Duff (2007)		3		MCI (8)		1
	Rasmusson (1995)	0.5			HE(45)		
BNT	*Flanagan (1997)*	0.91	0.25–0.5	(-2–2)	HE (31)	(s)	(0.37)
	Mitrishina (1995)	0.62-0.98	12 & 24		HE (122)	(ns)	
	Zec (2005)	0.76	9–15 & 18		HE (353 & 540)	0.19-0.21 (ns)	
MMSE	*Folstein (1975)*	0.83–0.89					
	Lezak (2004)	0.99	1		Dem		
	Tombaugh (2005)	0.65	3 & 60		HE (160)	(s) (ns)	0.5
COWAT (FAS)	*Harrison (2000)*	0.82	0.25–2		HE (90)	(s)	
	Snow (1988)	0.71	12		HE	(ns)	
	Duff (2007)	3			MCI (8)		-0.4
TMTA	*Spreen (1998)*	0.6–0.9					
	Duff (2007)		3		MCI		8.5 sec
TMTB	*Spreen (1998)*	0.6–0.9					
	Duff (2007)		3		MCI		-20 sec
	Basso (1999)		12			(ns)	
PAL	*Fowler (1995)*	0.64–0.88	1		HE;AD;QD		
RCFT copy	*Mitrushina (1991)*	0.56–0.68	4		HE		
	Meyers (1995)		6		HA (12)	(ns)	
RCFT immed	*Berry (1991)*	0.47	12		HE (41)		
	Meyers (1995)	0.76	6		HA (12)		
RCFT delay	*Berry (1991)*	0.59	12		HE (41)		
	Meyers (1995)	0.89	6		HA (12)		
Dual Task (mu)	*Rabbitt (1997)*	0.44			(33)		

RCI = Reliable Change Index; EF= effect size; CS = Change Score; HA = healthy adults; N = neurological; HE = healthy elderly; MCI = Mild Cognitive Impairment; AD = Alzheimer's Disease; Dem = Dementia; QD = Questionable Dementia; r = Pearson's correlation coefficient; n= number of participants

5.1.3 Methods

Sixteen of 24 healthy elderly control subjects agreed to return to complete the neuropsychological assessment battery on a second occasion. Parallel versions were available for two of the episodic memory measures comprising the research battery (HVLT-R and RCFT). Whilst no specific hypotheses were made regarding the predictive value of practice effects, a decision not to employ parallel test forms was taken in view of their potential added differential diagnostic or prognostic contribution (Schrijnemaekers et al. 2006) as well as the scarcity of data pertaining to form equivalence (Brandt 1991) and the relatively long intervals between assessments. Time intervals between first and second assessments ranged from 9 to 41 months with an average interval of just over two years (mean=28 months; SD=9.05) between test and re-test.

Details of the method of subject recruitment, participant characteristics together with psychometric and administrative characteristics of the cognitive screening measures employed can be found within Chapter 2, Materials and Methods, sections 2.1 and 2.

5.1.3.1 Statistical Analysis

Performances on each neuropsychological measure at baseline and follow-up were compared by way of paired t-tests, where data were normally distributed, and the Wilcoxon test, where visual inspection together with Kolmogorov-Smirnoff Z tests revealed that data were not normally distributed. Test re-test reliability was assessed using Pearson's correlations for parametric data and Spearman's rho for non-parametric data. Where differences in baseline and retest scores were found to be statistically significant, effect sizes were calculated for the magnitude of score change using the equation below, as a means of ascertaining the likely practical significance of the difference score.

[r = square root t squared / t squared + df] (Field 2005)

Where the effect size is denoted by r, t represents the t-value and df represents the degrees of freedom.

Lower and upper 95% limits of agreement were calculated for the mean interval score for each variable, by subtracting 2 * the standard error of measurement (SEM) from the mean score change (T1-T2) to obtain the lower limit value and adding 2 * the SEM to the mean score change to obtain the higher limit value, in accordance with the methods for assessing agreement between measurement proposed by Bland and Altman (Bland and Altman 1986).

As the limits of agreement are only estimates of the values which apply to the whole population (Bland and Altman 1986), 95% upper and lower boundary confidence limits of difference values were then calculated for each variable (by subtracting 2 * SD of the mean score change for each variable from the lower boundary limit value and adding 2 * SD of the mean score change to the higher boundary limit value), such that fewer than 2.5% of any given sample of healthy elderly would be expected to show a change of a greater magnitude in performance than that denoted by the upper confidence limit value over the average 28 month period.

Where change score distributions were not normally distributed (i.e. in the case of the immediate and delayed recall components of the ACE, the HVLT-R DI, the EENT and the copy component of the RCFT), upper and lower quartile range scores that most closely corresponded to the lowest and highest 5th%ile of change scores were obtained and compared to the boundary cut off points derived as above.

For the immediate and delayed recall subtests of the ACE, reliable change index scores calculated in this manner were less conservative than those calculated based on assumptions of normality. For the variables HVLT-DI, RCFT copy & PAL 6 box errors, variability in the reliable change indices obtained by these two methods were reflective of one or more outlying values. When the quartile ranges for HVLT-DI and RCFT copy change scores were re-examined following omission of their respective outliers, similar change index scores to those calculated based on the assumption of normal distributions were obtained (HVLT-DI = (-3 – 0); RCFT copy = (-3 - 1)).

For the purposes of subsequent analyses (see chapter 6), MCI participants who exhibited declines in performance falling outwith the upper confidence limit set for each neuropsychological variable, were considered to have undergone clinically significant levels of decline in test performance. Conversely, improvements in performance of a magnitude greater than the upper set confidence limit were taken to a clinically significant level of improvement.

All comparisons are based on a sample of 16 healthy elderly control subjects with the exception of the GNT, where a baseline score was missing for one of the healthy control participants.

5.1.4 Results

In Table 5.2 it can be seen that the control group performed at significantly higher levels at re-test as compared to baseline on the ACE total ($t(15) = -2.24$, $p < .05$), HVLT-R total ($t(15) = -2.68$, $p < .05$) and Rey Delayed recall ($t(15) = -2.44$, $p < .05$), respectively. By contrast, significant declines in the mean performance of the control group on the Category Fluency ($t(15) = 2.61$, $p < .05$) and Part A of the Trail Making Test ($t(15) = -2.69$, $p < .05$) were noted at re-test relative to baseline. Where statistically significant differences in baseline and follow-up scores were observed, associated effect sizes were uniformly large (i.e. > 0.5;(Field 2005), indicating the likely practical significance of these difference findings. The mean performances of the control group at baseline and follow-up did not differ significantly on the remaining neuropsychological measures (in all cases $p > .05$).

Re-test reliabilities for the control group were for the most part large (i.e. > .5), with the exception of the MMSE, the immediate recall sub component of the ACE, PAL errors at the 6 box stage, the Discrimination Index from the HVLT-R, the paper and pencil Dual Task and the copy component of the RCFT, where correlations were of a medium (i.e. r= .3 – .5; in the case of the MMSE) and otherwise small (r =.3) magnitude.

The magnitude of score change required to surpass the upper 95% confidence limit ranged from a minimum of 1 word for the total recall measure of the HVLT-R to a maximum of 4 items for the BNT. Conversely, with reference to lower 95% confidence boundaries, a

Table 5.2 Comparison of healthy elderly control performances on neuropsychological measures at baseline and follow-up

Measure	Baseline Mean (SE) ^	Re-test Mean (SE) ^	Mean change score (SE)	95% Conf limits diff values	5th %ile quartiles diff values	T(df)	ES	Test-retest rel
ACE								
Total	94.63(0.86)	96.13(0.72)*	1.50(0.67)	-2.70 – 1.71		-2.24	0.71	.65***
Immed a	21.00(17-21)	21.00(19-21) ns	-0.31(0.33)	-4.35 – 3.59	-4 – 1	16.0		-37ns
Delay a	6.50 (5-7)	7.00(5-7) ns	-0.13(0.18)	-2.00 – 1.19	-2 - 0	5.00		.73**
MMSE a	29.00 (28-30)	29.50(27-30) ns	0.06(0.30)	-2.87 – 2.99	-2 – 1	27.0		.39ns
HVLT-R								
Total	23.88(1.43)	27.13(1.20)*	-3.25(1.21)	-0.47 – 0.64		-2.68	0.57	.59*
Delay	8.19(0.68)	9.00(0.56) ns	-0.81(0.52)	-1.78 – 2.45		-1.57		.67**
DI a	10.50(5-12)	11.00(9-12) ns	-0.63(0.52)	-2.99 – 3.80	-7 – 0 (-3 – 0)^	33.0		.28ns
PAL a	8.00(1-30)	6.5(0-46) ns	0.56(3.11)	-3.39 – 3.52	-32 – 12 (-9 -9)^	59.0		.12ns
RCFT								
Copy a	35.50(25-36)	36.00(34-36) ns	-1.25(0.73)	-5.01 – 3.50	-11 – 1 (-3 – 1)^	6.50		.28ns
Immed	20.47(1.45)	23.28(1.99) ns	-2.81(1.45)	-2.46 – 1.66		-1.95		.69**
Delay	19.25(1.80)	22.16(2.13)*	-2.91(1.19)	-2.00 – 1.19		-2.44	0.53	.83***
BNT a	58.50(48-60)	58.00(50-60) ns	0.44(0.51)	-3.66 – 3.98	-2 – 3	37.0		.49ns
GNT	24.40(0.80)	23.60(1.00) ns	0.80(0.68)	-4.53 – 2.87		1.18		.78**
GFT	21.06(1.03)	20.25(1.22) ns	0.81(0.94)	-2.01 – 2.39		46.0		.67**
EENT	48.44(1.02)	48.00(0.63) ns	0.44(0.72)	-1.82 – 2.06		43.0		.72**
C-Flu	53.69(2.91)	49.13(3.42)*	4.56(1.75)	-1.01 – 1.73		2.61	0.56	.86***
TMT								
Part A	38.06(2.28)	43.81(2.54)*	-5.75(2.14)	-2.78 – 1.52		-2.69	0.57	.61*
Part B a	80.00(43-125)	83.50(47-216) ns	-8.56(8.72)	-2.62 – 2.15	-19 – 5	39.0		.53*
Dual	92.34(3.35)	98.02(2.11) ns	-5.68(3.81)	-4.28 – 3.17		-1.49		.08ns
FAS	50.38(3.45)	50.06(4.58) ns	0.31(2.56)	-1.54 – 1.57		0.12		.83***

*** = p<.001, ** = p<.01, * = p<.05, (ns) = non-significant; ACE, Addenbrookes Cognitive Examination; Immed, immediate recall of name and address; Delay, delayed recall of name and address; Conf, confidence; Diff, difference; HVLT-R, Hopkins Verbal Learning Test – Revised; Total, total words recalled; Delay, words recalled at delay; DI, discrimination index; PAL, CANTAB Paired Associated Learning; RCFT, Rey Complex Figure Test; BNT, Boston Naming Test; GNT, Graded Naming Test; GFT, Graded Faces Test; EENT, Edinburgh Exemplar Naming Test; C-Flu, total number of animals, fruits and vegetables; TMT, Trail Making Test; FAS, controlled oral word association test; a, Data analysed non-parametrically; ^, the median and range are given; ^ upper and lower quartile ranges most closely corresponding to 95%ile cut off points following removal of outliers; EF, effect size; rel, reliability

minimum score change of -1 word was required for the total recall measure of the HVLT-R and a maximum of -6 points for the copy component of the RCFT.

The lowest ACE value obtained for the control group at retest was 92/100, indicating the absence of any control participants who had developed a dementing illness during the course of the study

5.1.5 Discussion

The majority of neuropsychological measures in our study battery (i.e. 14/20) demonstrated adequate levels of re-test reliability (i.e. $r > .5$) amongst healthy elderly participants across time intervals ranging from 9 – 41 months, attesting to their general suitability of use in monitoring cognitive ability across time.

The significant improvements in the follow-up performances of the control group on the ACE, RCFT delayed recall and HVLT-R total recall scores and the large effect sizes associated with the mean score changes warrant further discussion. Practice effects in performance on the recall component of the HVLT were previously reported within a sample of healthy elderly (Schrijnemaekers et al., 2006). The effects were not, however, observed within MCI and early AD samples.

Several authors have postulated that the absence of practice effects in MCI and AD patient's may be indicative of underlying disease pathology interfering with the ability of such patients to acquire and retain strategies to assist recall of new information across time (Darby et al. 2002;Schrijnemaekers et al. 2006) It has consequently been suggested that observed lack of practice effects may be of potential utility in the identification of MCI and the early and differential diagnosis of AD. The relatively small reliable changes indices observed in association with this sample of healthy elderly would suggest that the learning component of the HVLT-R may be especially sensitive to small changes in verbal learning ability that are of a clinically significant magnitude.

Alternate forms were not used in the present study, in view of the potential added predictive value of practice effects (or lack thereof), the relatively large i.e. 12 month re-test intervals and the scarcity of data attesting to form equivalence (Brandt 1991). Our findings would suggest that this may be a necessity if one wishes to avoid taking into account effects of

practice, even across relatively lengthy intervals. It is possible that the structured nature of the HVLT-R (i.e. with words belonging to one of four categories) makes it more susceptible to practice effects than other popular verbal list learning tasks i.e. RAVLT, with retention of, or familiarity with the semantic categories serving to enhance recall performance at re-test. Alternatively, some of the variability in re-test scores may be accountable for in terms of variability in the administration styles (i.e. clarity of speech and/ or time given to respond) of baseline and re-test examiners.

Significant practice effects were also observed across an average 28-month period on the ACE, within our sample of healthy elderly control participants. These findings are of considerable clinical importance, as re-test data for the ACE is not available despite its widespread use in clinical practice as a means of monitoring cognitive deterioration and response to drug therapy in early AD and MCI patients. It is possible that some of this variation stems from the differing administration or scoring styles of the baseline and re-test examiners, although reference was made, by all examiners, to the published scoring and administration guidelines for the ACE in an attempt to minimise this risk.

Our findings imply that failure to take into account practice effects on this popular bedside cognitive screening measure may lead to underestimation of the magnitude and clinical significance of small declines in total ACE scores. There is also the risk of falsely attributing improvements in performance on the ACE to treatment effects. Although alternate forms have been developed for the revised version of the ACE i.e. ACE-R (Mioshi et al. 2006), information about form equivalence is lacking and it is unclear to what extent clinicians make use of the alternate forms. Furthermore, whilst details of the name and address for recall vary between each version of the ACE-R, the remaining subtests are unchanged. For these reasons, our preliminary findings using the ACE suggest that replication in a larger sample of healthy elderly with the more recently devised ACE-R is warranted.

Significant practice effects in the order of an additional 2 points, were also observed for the delayed recall component of the RCFT. The significance of practice effects has not been reported in previous research, although the potential for contamination of the incidental nature of the learning paradigm to affect re-test recall performance has been noted and it is

possible that even across the relatively lengthy re-test intervals comprising the present study, the loss of the 'incidental' learning and novelty components of this task affected recall at follow-up. In view of the minimal instruction accompanying this task, one would not expect to observe large differences in an individual's recall ability as a function of different examiners. Test re-test correlations for the RCFT in our sample of healthy elderly mirrored those of previous studies where the lowest correlation values (likely attributable to restriction of range in healthy elderly samples) have generally been reported for the copy component (Berry et al. 1991;Meyers and Meyers 1995), and higher correlation coefficients are reported in association with delayed recall (Meyers and Meyers 1995;Mitrushina and Satz 1995).

A significant increase was observed in the average time taken to complete Part A of the Trail Making Test at follow-up. This finding is in keeping with a body of previous research work documenting a decline in psychomotor processing speed as a function of age (Lezak 2004) as well as with the broader view that declines in processing speed account for 'much if not all of the measured changes in performance that occur with age' (Salthouse 1996). As the instructions, task and scoring requirements for Part A of the TMT are relatively straight forward, with no (or minimal) on-task examinee - examiner interaction, there is little reason to expect to observe low levels of inter-rater reliability for this task. The apparent sensitivity of measures of psychomotor processing speed to the effects of normal ageing within this and other studies implies that Part A of the Trail Making Test may lack the specificity required to differentiate between normal ageing and early dementia or track small but clinically significant declines in the cognitive function of individual patients across time.

Significant declines in performance, across an average 28 month period, were also observed on the Category Fluency task. This finding was unexpected and at odds with previous findings of significant practice effects, in the order of +1.3 words, among 106 'normal' adults with a mean age of 57 years (Bird et al. 2004). It is conceivable that the relatively short i.e. one month re-test interval for the former study facilitated practice effects whilst simultaneously eradicating the effects of cognitive aging on lexical retrieval. Several studies have documented a decline in the number of animals generated within a 60 second time interval with advancing age (Fama et al. 1998;Lucas et al. 1998). There is some room for variability in scoring this task, with regards to the extent of 'subcategorisation' that is tolerated by the examiner (i.e. fish, salmon, bream, flathead, cod). Furthermore, the extent to which the examinee feels at ease within the testing environment could conceivably influence his/her strategy and verbal generative capacity. As such, it is plausible that some the the re-

test variability observed for this task is accountable for in terms of the use of different examiners at baseline and follow-up. The reliability coefficient for the category fluency task using animals, fruits and vegetables (r = .86) was a little higher than that previously reported (r=.56; (Bird et al. 2004), which is perhaps not surprising in view of the inclusion of 3 categories, as opposed to just one, in the present study.

The lowest test re-test reliability value was reported in association with the combined dual task score from Sergio Della Sala's Dual Performance Task (Della Sala, personal communication, 2008; (Della Sala 2005). Previous research findings have documented similarly poor test re-test reliability (Baddeley et al. 1997). In keeping with previous findings, much of the instability in dual task performance amongst our healthy elderly control group stemmed from the digit span (r=.27) as opposed to the tracking (r=.48) component of this task. It is plausible that the low test re-test reliability observed for the digit span component of this task was compounded by the use of different raters at baseline and re-test. This is problematic within the context of early detection of AD where both the sensitivity and reliability of neuropsychological measures needs to be high in order to detect clinically meaningful deficits and declines.

The reliability coefficient for the number of errors made at the 6 box stage of the PAL subtest was disappointingly low (r= .12) and less favourable than corresponding values previously reported in association with total error numbers (Fowler et al. 1995). The average age of healthy control participants (59 years), and the interval between re-test (i.e. 1 month) were considerably lower, in the latter study. These factors together with the fact that the error score used in our study was derived from performance across a restricted number of trials (i.e. during the 6 box learning trials only), and examiners differed at follow-up, may account for some of the discrepancy in study findings.

The similarly low re-test reliabilities observed for the MMSE, the discrimination index score from the HVLT-R and the immediate recall component of the ACE likely reflect the restricted range of scores, (i.e. MMSE at T1 28 min - 30 max; ACE immediate recall at T2 19 min -21 max; HVLT-R DI at T2 9 min – 12 max) that was observed in association with each of these measures.

5.1.6 Conclusion

In summary, reliable test performances were observed on a majority of neuropsychological tasks across a mean 28 month interval in our healthy elderly control group. Practice effects were present for several episodic memory measures together with a widely used cognitive screening measure (ACE) and these should be taken into account when interpreting change in performance on such measures over time. The ceiling effects (MMSE, ACE immediate recall, HVLT-DI, RCFT copy) and low re-test reliabilities (PAL, Dual Task) observed for several other tasks imply these measures may be less suited to monitoring cognitive performance and detecting small but significant changes in performance across time, in patients with mild cognitive difficulties. The significant overall decline in the performance of the control group on several timed tasks cautions against the use of processing speed measures for monitoring purposes in an elderly population, without reference to reliable change index scores.

6. Neuropsychological predictors of Outcome in MCI

6.1.1 Abstract

Background

Cognitive impairment is known to predate the point at which a diagnosis of AD can be made on clinical grounds. Neuropsychological performance in the years prior to diagnosis may therefore help to predict whether or not a patient with cognitive complaints will develop dementia. The case definition of aMCI, within existing neuropsychological longitudinal follow-up studies, varies on a number of important levels, giving rise to disparity in reported annual rates of conversion from aMCI to dementia and complicating comparison of findings from different studies. There is little information of an empirical nature addressing the impact that variability in the use of cognitive measures, psychometric cut-offs, and requirement(s) for single or multiple same domain impairment, has on risk of conversion from aMCI to dementia. If the predictive validities of one or more neuropsychological measures were replicable, they could be used as part of a wider cognitive evaluation within specialist memory clinic settings, to provide important information of a differential diagnostic and prognostic nature.

Aims

The study aimed to determine 1) the annual rate of conversion from aMCI to dementia 2) whether (and which) neuropsychological performance(s) at baseline could be used to distinguish between aMCI patients who would and would not go on to develop dementia 3) the robustness of previous longitudinal findings using the ACE, GNT and PAL in predicting conversion from aMCI to dementia, within a larger sample of aMCI patients with higher general level of cognitive functioning at baseline and across a lengthier period of follow-up, and 4) the levels of accuracy with which a combination of neuropsychological measures and clinical/demographic information could predict the fate of our aMCI cohort.

Method

46 patients with aMCI who underwent extensive neuropsychological assessment at baseline and annually thereafter, were followed up for an average of 4 years and classified as having developed dementia or not in accordance with the presence/absence of a clinical diagnosis of dementia recorded in their medical file at the study endpoint. Differences in baseline cognitive performances of converters and non-converters were analysed using independent

sample t-tests, and levels of classification accuracy were determined by way of sensitivity, specificity, positive and negative predictive values, ROC analysis and logistic regression.

Results

Forty one percent (18/44) of the aMCI participants had received a clinical diagnosis of dementia (mostly AD) by the study endpoint, yielding an annual conversion rate of 10% within our aMCI clinic sample. Performance on 2 neuropsychological measures at baseline (but none of the demographic indices) differentiated converters from non-converters. Using a combination of ACE and/or PAL the overall rate of classification accuracy 68% was lower than previously reported. A regression model containing both ACE total score and HVLT-R DI score yielded an overall classification accuracy (for aMCI converters vs. aMCI non-converters) of 74%.

Conclusion

A significant proportion of patients fulfilling criteria for aMCI go on to receive a clinical diagnosis of dementia. The vast majority of aMCI patients within our sample displayed persisting or progressive impairment of a cognitive nature over the course of the study. Performance on two neuropsychological measures at baseline was helpful in discriminating between future aMCI converters and non-converters beyond demographic information. Bedside cognitive screening scores falling above cut off point for dementia are not necessarily commensurate with the absence of a pre-clinical dementia. The pervasiveness of the memory impairment in aMCI together with the high risk of conversion to dementia and predictive validity of several cognitive measures has implications for the clinical management of patients with aMCI.

6.1.2 Introduction

Criteria for aMCI were devised in an attempt to capture and define the pre-clinical phase of AD. There is evidence to suggest, however, that in its current form, aMCI comprises a heterogeneous group of patients (Blossom et al. 2007;Chertkow et al. 2007), some of whom will progress to dementia with time and others who will not (Visser et al. 2006). The clinical usefulness of MCI as a diagnostic label has been challenged on these grounds (Petersen 2005b) . To maximise the diagnostic value of aMCI, criteria must give rise to a more homogenous group of pre-clinical dementia sufferers. To this end, criteria should be applied in a manner that reflects our current knowledge of the predictive value of specific neuropsychological measures and their combinations. The application of MCI criteria in clinic and population settings is known to give rise to different subsets of patients who show different rates of conversion to dementia over time (Chertkow et al. 2007). Consideration of which cognitive measures to employ in the assessment of aMCI should therefore be context specific.

A number of clinic-based longitudinal studies have sought to characterise the neuropsychological profiles of aMCI patients who do and do not go on to develop dementia, and to determine the predictive value of performance on neuropsychological measures among patients with aMCI. A summary of study characteristics and primary findings can be found in Table 6.7. There is considerable variability in the length of follow-up, the general levels of cognitive functioning of the aMCI cohorts at baseline (as indicated by mean MMSE scores), the mean ages of the aMCI cohorts, and the neuropsychological measures and cut off levels for impairment that are employed to define aMCI.

Even across studies that are clinic based, it is conceivable that such variability could influence the annual rate of conversion from aMCI to dementia that is observed. As the prodromal phase of AD is likely to extent beyond a two-year period (Amieva et al. 2005) a substantial number of these studies with shorter follow-up periods risk false assignment of MCI non-converters. Furthermore, in the case of shorter follow-up periods, one might expect the baseline neuropsychological performances of the preclinical dementia participants to show greater impairment than would be present earlier on (i.e. 4 years before onset) in the course of their illness. This would in turn have the effect of increasing the magnitude of predictive validities associated with baseline performance on neuropsychological tasks. A

relatively small number (i.e. 2/14) of existing clinic based aMCI longitudinal studies report follow-up periods extending beyond 3 years (Fox et al. 1998;Tabert et al. 2006). For one of these, the aMCI cohort comprised participants of a younger age with an established family history of AD who were asymptomatic at baseline, complicating comparison with other non-familiar, symptomatic 'older' aMCI cohorts (Fox et al. 1998).

Similarly, where cognitive screening scores (at study entry) fall below the commonly applied MMSE cut off for dementia (i.e. 27/30), one might also expect to see greater levels of baseline neuropsychological impairment, which would facilitate the discrimination of future MCI converters and non-converters. To this end it should be noted that in just 3 of the 14 clinic based longitudinal studies identified, (where the relevant data is provided), the baseline mean MMSE score for the MCI converter group falls above the higher cut off for dementia (i.e. 27/30). Levels of educational attainment for the aMCI participants were in all cases average at least, suggesting that an underlying dementia process might well have been suspected on the basis of even the most rudimentary of cognitive screening instruments alone, leading one to question the requirement for fuller cognitive evaluation.

The comprehensiveness of the neuropsychological battery employed and the appropriateness (on both theoretic and empirical grounds) of test selection might also be expected to influence predictive validity findings. Whilst there is a good deal of variability in tests employed across existing clinic studies, several measures included in our own research battery (i.e. MMSE, ACE, semantic fluency, COWAT, BNT, famous face naming, PAL, TMTB and GNT) have been investigated previously, and found to differentiate MCI future converters from non-converters (Ahmed 2008b; Albert 2001; Amieva 2004; Blackwell 2004; Estevez-Gonzalex 2004; Griffith 2006; Lee 2006; Rami 2007; Tabert 2006; Thompson 2002).

Consistent across all studies is the finding of one or more measures of episodic memory function as a significant predictor of future conversion to dementia among MCI patients. There is nonetheless a good deal of variability in terms of the precise measure of episodic memory function chosen and the nature of the memory paradigm. Furthermore, where additional non-memory measures have been reported as significant co-predictors, these have included measures of processing speed (Albert et al. 2001;Tabert et al. 2006), and object

naming tasks (Blackwell et al. 2004). High levels of discrimination between aMCI and normal controls have been reported in association with composite scores derived from performances across a range of memory and non-memory tasks (Loewenstein et al. 2007b). Despite this, only one clinic-based study of MCI patients has investigated the predictive validity of baseline composite scores for neuropsychological tasks. Perri et al (Perri et al. 2007) reported a sensitivity of 75% and a specificity of 68% for conversion from MCI to dementia across a two-year period using a cumulative delayed recall index score.

The highest levels of overall classification accuracy (i.e. 100%) are reported over a 2.5 year interval for a combination of the PAL, age and GNT (Blackwell et al. 2004). High levels of negative predictive validity (i.e. 100%) have also been reported using the PAL in combination with the ACE (Ahmed et al. 2008b) and the GFNT (Thompson et al. 2002) across shorter i.e. 1 year intervals. These findings have not been replicated outside the test authors group, in larger numbers of aMCI patients, across follow-up periods extending beyond 2 ½ years and where the mean general level of cognitive functioning falls above cut off points for dementia. If replicable, however, these measures could be used in the neuropsychological assessment of aMCI within specialist memory clinic settings, to provide information of a differential diagnostic and prognostic nature to this patient group. Clinic based studies using combinations of other neuropsychological variables report overall classification accuracies (i.e. aMCI converter vs. aMCI non-converter) in the range of 72.4% - 86%.

The present study represents the longest detailed neuropsychological aMCI clinic-based follow-up study, with an average of 4 years from baseline neuropsychological assessment until final review of the medical notes. It is the only study with a follow-up period extending beyond 3 years, for which baseline cognitive screening scores for the aMCI converter group fall above the higher level cut off point for dementia (i.e. 27/30 MMSE), and is one of two clinic based studies to examine the utility of composite cognitive scores at baseline as predictors of conversion to dementia in MCI patients. Furthermore, to the best our knowledge, this study represents the first clinic based study to investigate the robustness of the, GNT, GFNT and a combination of the ACE and PAL as predictors of conversion from aMCI to dementia outwith the authors' research group, and to report the detailed fate of aMCI non-converters in terms of the course of their cognitive impairment.

6.1.3 Methods

Forty six patients who fulfilled criteria for aMCI and undertook an extensive battery of neuropsychological measures at baseline were followed-up over an average of 4 years. At the end of this period aMCI participants were grouped in accordance with whether or not they had received a clinical diagnosis of dementia (as documented in their medical file) at any point subsequent to their initial study assessment. Patients were also classified in accordance with whether or not they met our own study criteria for dementia, based on their performance on the neuropsychological assessment battery and the level of functional impairment endorsed by a significant other at their final assessment session, together with the magnitude of cognitive decline in domains other than that of episodic memory, across their years of study participation (for details see Chapter 2, Materials and Methods section).

6.1.3.1 Materials, Participants & Outcome Criteria

For details of test materials and participants see Chapter 2, Materials and Methods, sections 2.1 & 2.2. The primary outcome was whether or not the aMCI participant had received a diagnosis of dementia at study endpoint. Dementia diagnosis at study endpoint was determined using two methods, 1) whether or not a diagnosis of dementia had been reached on clinical grounds and recorded in the participant's notes, and 2) in accordance with a set of study criteria that allowed for the documentation of significant levels of cognitive impairment, progressive cognitive decline and functional impairment at study endpoint. Criteria for the two methods of determining outcome are summarised in Table 6.1 below.

Table 6.1 Criterion for determining an outcome of dementia at study endpoint

Medical Notes	Study Criteria
Clinical diagnosis of dementia documented in the medical file at any point following the participant's initial (i.e. baseline) study assessment	A performance of 1.5 SD or more below study control mean on two or more episodic memory measures **AND/OR** 1.5 SD or more below control mean on two or more non episodic memory measures **AND** Evidence of significant declines in non-memory domains on two or more tests of a cognitive domain other than that of (or in conjunction with) memory (as defined by a decline of a magnitude that would be expected to occur in less that 2.5% of a normative sample over the course of the study. **AND** Evidence of significant functional decline (as defined by 3 or more ratings of 3 or 4 in 2 or more non-memory domains on the MCI ADL scale (Farias et al. 2006) **OR** Evidence of decline in two or more areas of the IADL scale (Lawton and Brody 1969)

6.1.3.2 Statistical Analysis

Independent t-tests were conducted to compare the baseline performances of aMCI converters and aMCI non-converters on the demographic indices of age, NART FSIQ, and years of follow-up, and the neuropsychological measures; ACE total score, PAL, HVLT-Delayed recall and DI, GFT, Category fluency and TMTB. The alpha level was adjusted to control for multiple comparisons using Holm's sequential Bonferroni correction method (Holm 1979) wherein the p-value of the largest effect size was compared to alpha (i.e. 0.05) divided by the number of t-tests (i.e. 10). If this comparison was significant, then the p-value of the next largest effect size was compared to alpha (i.e. 0.05) divided by the remaining number of comparisons (i.e. 9), and so on until the p-value of the 10^{th} test was compared to alpha 0.05.

The demographic indices were compared on account of their established influence on risk of developing late onset dementia (Mebane-Sims 2009;Petersen et al. 1999;Whalley et al. 2000). Seven neuropsychological measures (ACE, PAL, HVLT-R DI & delayed recall,

GNT, GFT, TMTB) were selected from a total of 18 on the basis of 1) their high levels of sensitivity to aMCI relative to other measures of similar cognitive functions as established via our own cross-sectional analyses or 2) high levels of predictive validity as established by one or more clinic based longitudinal study of neuropsychological predictors of dementia.

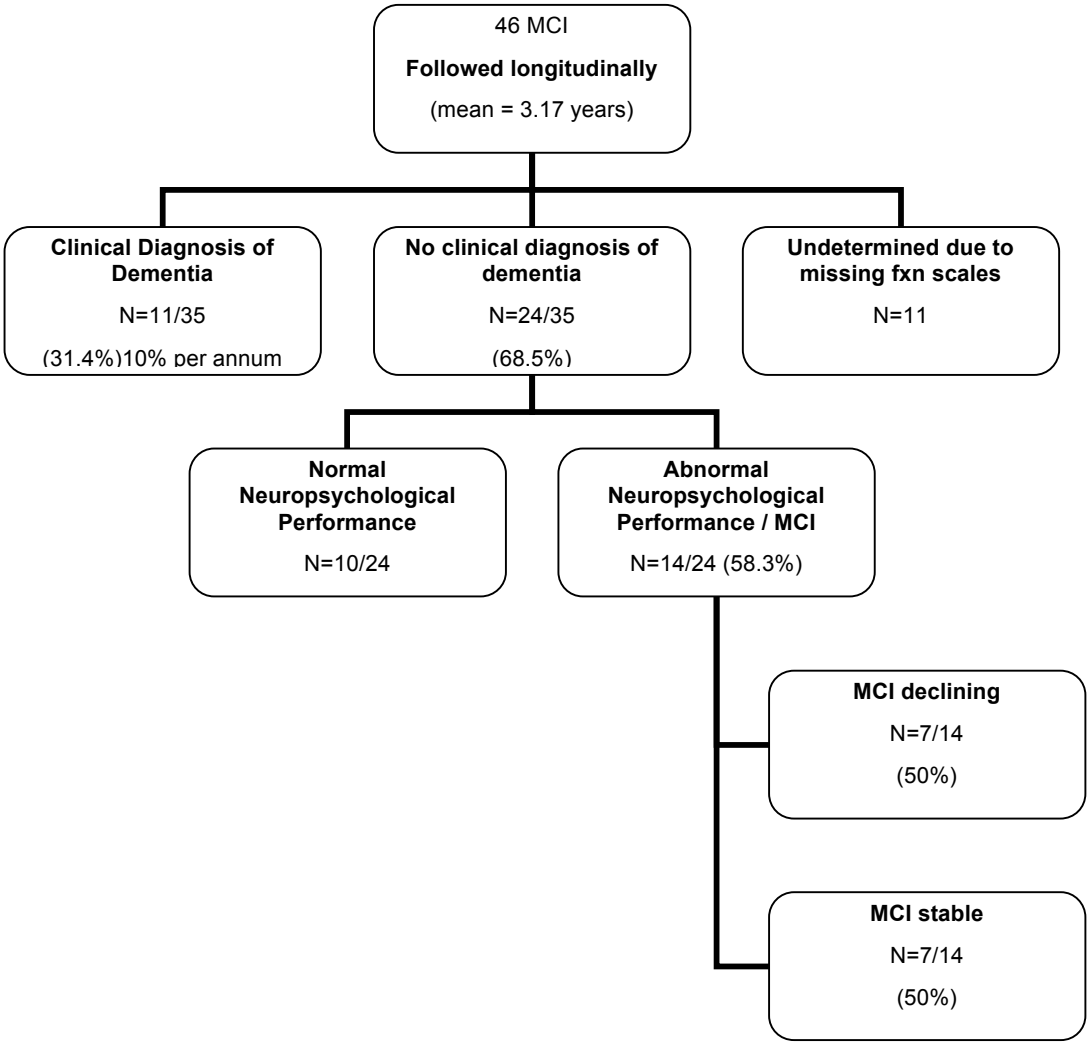
The percentages of aMCI patients who fulfilled criteria for dementia at the study endpoint were calculated in accordance with medical file and study criteria. AMCI patients who did not meet study criteria for dementia at their final assessment were further classified in accordance with percentages exhibiting a normal vs. persisting aMCI neuropsychological profile, and percentages showing a progressive, as compared to a non-progressive cognitive course during their years of study participation. Where outcomes (i.e. dementia or no dementia), could be reached in accordance with both the medical file and study criteria, they were cross tabulated to determine levels of agreement between these two methods.

Sensitivity, specificity, positive and negative predictive values, together with the overall percentage of classification accuracy in predicting conversion or non-conversion to dementia in our aMCI cohort, was determined using a combination of a total score of <88/100 on the ACE or a performance of 2 SD or more below controls on the PAL, as this combination of measures has previously been associated with perfect i.e. 100% sensitivity and negative predictive value (Ahmed et al. 2008b).

Neuropsychological measures for which the baseline performances of aMCI converters and aMCI non-converters were significantly different, were entered simultaneously alongside the demographic variables of age, NART FSIQ and years of follow-up into a logistic regression analysis. The FSIQ score for one of the aMCI participants, who received a clinical diagnosis of dementia during the course of the study, was missing, and the corresponding data for this participant are for this reason not represented in the regression analysis and final model. A backward stepwise procedure using the likelihood ratio was applied to determine model content and levels of overall classification accuracy. Criterion for entry and removal were set at $p=0.05$ & $p=0.01$ respectively, with 20 iteration. The Likelihood method was chosen in order to minimise the risk of falsely rejecting a predictor that is a significant contributor to outcome, as can occur where regression coefficients are large and the Wald statistic is employed as a means of ascertaining a variables predictive worth (Field 2005).

6.1.4 Results

Figure 6.1 Flowchart of aMCI endpoint classification in accordance with study outcome criteria

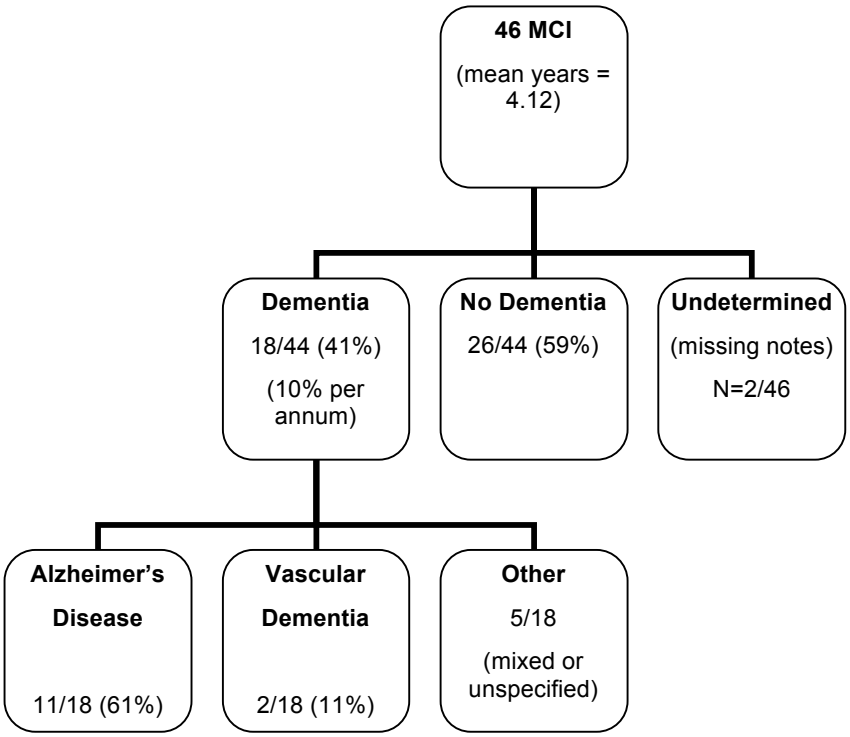


Normal Neuropsychological performance is defined by a performance of 1.5sd or more below control mean on one or less neuropsychological task at study endpoint; Abnormal neuropsychological performance or MCI is defined by a performance of 1.5sd or more below control mean on two or more neuropsychological tasks at study endpoint (chance rate among the 22 MCI participants without a diagnosis of dementia = 0.83 i.e. less than 1 person). MCI declining is defined by deterioration of a magnitude seen in fewer than 2.5% of any given sample of healthy elderly over an average 28 month period in at least two measures from either the domain of semantic memory or executive functioning.

In figure 6.1 it can be seen that of the 35/46 aMCI participants where data was complete at their final follow-up, 11/35 or 31% met study criteria for dementia. Eighteen of 35, or 51%,

either fulfilled study criteria for dementia at their final assessment and/or continued to decline cognitively over the course of their study participation. Of those 17/35 aMCI participants who did not display evidence of progressive cognitive decline over the course of the study, 7 displayed persisting cognitive impairment of a non-progressive nature, 10 showed resolution of their cognitive difficulties by their final study assessment. Four of the 10 who showed 'resolution' of their cognitive difficulties at their final study assessment met criteria for aMCI at the point of study entry (as established via performance on baseline clinical assessment) but not at their first (i.e. baseline) study assessment session. Cognitive impairment in keeping with dementia or MCI remained present in 71.4% of the aMCI cohort over the average 3-year follow-up period.

Figure 6.2 Flowchart of aMCI endpoint classification in accordance with clinical diagnoses in medical notes



Forty one percent of aMCI participants in our study received a clinical diagnosis of dementia at some point prior to the study endpoint, yielding an annual conversion rate of 10% per annum. This figure rose to 47% following exclusion of the 6 MCI participants who were

recruited into the study on grounds of filling aMCI criteria at their clinical assessment, but were no longer showing sufficient evidence of impairment to meet aMCI criteria at their baseline study assessment. In 61% of such cases, the diagnosis reached was that of AD. In a remaining 30%, (where specified), the diagnosis was of VD or mixed VD / AD. Fifty nine percent of aMCI patients in our sample had not received a clinical diagnosis of dementia an average of 4 years post their initial study assessment.

Table 6.2 Agreement in final outcome for aMCI participants in accordance with study and medical file criteria determining dementia diagnosis

Count

		Dementia Diagnosis Medical Notes		
		no dementia	dementia	Total
Dementia diagnos study criteria	no dementia	19	4	23
	dementia	4	6	10
Total		23	10	33

Outcome criteria were available for both study and medical file methods for 33/46 aMCI participants. For these 33 participants, there was 76% agreement in the two methods of diagnoses i.e. in 25/33 cases the diagnosis reached was the same for both methods. For the 8/33 cases where there was a disagreement in final diagnosis according to the two different methods, in half i.e. 4 of these cases, a diagnosis of dementia was reached via medical files but not our study criteria. In each such case, the diagnosis of dementia was reached at least one year post the aMCI participant's final study assessment.

In the other 4 cases of disagreement, a diagnosis of dementia was reached via our study criteria but not the medical files. In two such cases, the attending Consultant Psychiatrist was notoriously late in establishing clinical diagnoses of dementia. In a further case, there were significant medical co-morbidities, which may have conceivably complicated the process of reaching a dementia diagnosis. In the final case, it was not possible to ascertain any clear reasons for the disparity in classification at outcome across the study and medical file methods of determination.

Of the 24 aMCI participants who did not meet study criteria for dementia at the study endpoint, 14 fulfilled criteria for aMCI and 10 were classified as normal (based on a performance of 1.5 SD or more below normative control group on one or less of the 18 neuropsychological measures). Four of the 10 participants who were classified as 'normal' were performing more than 1.5 SD below the healthy elderly control mean on one non-memory measure of neuropsychological functioning only; 2 on one memory measure only and 4 were performing within 1.5 SD of the control mean on all neuropsychological measures. The latter 4 participants were recruited into the study on the basis of their impaired performance on memory measures during a clinical assessment but were no longer displaying deficits at their baseline study assessment.

If one were to assume that the neuropsychological variables were perfectly correlated i.e. r=1, then the chances of MCI participants showing impairment on two of these would be 0.07 or 7% (i.e. 3 MCI participants would do this by chance). If the 18 neuropsychological variables were all completely unrelated, i.e. r=0 then the chances of scoring 1.5 or more below the normal control group on two of these would be (0.07 * 0.07) = 0.005 (i.e. 0 MCI participants would do this by chance). If we are to assume that the average level of correlation between the 18 neuropsychological variables falls somewhere between r=0 and r=1, say at r=0.5, then chance alone could lead to 1 MCI participant performing 1.5sd or more below the normative control group on 2 of these neuropsychological variables, which represents a very low false positive rate. In the interest of adopting a conservative approach to defining persisting cognitive impairment, the criteria for persisting MCI at the study endpoint was specified as a performance of 1.5 SD or more below control group norms on 2 or more of the 18 neuropsychological measures.

Of the 14 MCI participants who did not receive a diagnosis of dementia according to study criteria, but for whom the endpoint neuropsychological assessment was abnormal, 7 displayed evidence of significant progressive cognitive decline over the course of their study participation, whilst the remaining 7 did not (where significant decline was defined by a deterioration in at least two measures other than that of memory i.e. either the domain of semantic memory or executive functioning, of a magnitude seen in fewer than 2.5% of any given sample of healthy elderly over an average 28 month period).

Table 6.3 Comparison of demographic indices and baseline neuropsychological performance of aMCI converter and aMCI non-converter groups

Variable	aMCI Converter Mean (SE)	aMCI Non-conv Mean (SE)	T	df	p-value	Effect size
Demographic Info						
Age	76.00 (1.59)	73.19 (5.40)	-1.53	42	0.134	0.23
NART IQ	116.41 (2.01)	117.42 (1.33)	0.44	41	0.664	0.07
Months of follow-up	51.39 (3.01)	52.35 (2.71)	0.23	42	0.818	0.04
Cognitive Screening						
ACE total	86.61 (1.32)	91.27 (0.93)	2.98	42	0.006* ~	0.42
Episodic Memory						
PAL 6 box errors	23.22 (3.62)	13.46 (2.45)	-2.32	42	0.026*	0.34
HVLT-R delayed recall	3.50 (0.70)	5.84 (0.67)	2.38	41	0.022*	0.35
HVLT-R DI	7.17 (0.67)	9.24 (0.40)	2.81	41	0.008* ~	0.40
Semantic Memory						
GNT	20.25 (1.16)	21.00 (0.76)	.56	39	0.582	0.09
GFNT	15.28 (1.34)	18.08 (0.86)	1.84	41	0.074	0.28
Attention/Executive						
TMT B +	152.72 (19.62)	100.92 (9.76)	-2.58	42	0.014*	0.39

+results were replicated using nonparametric equivalent analysis i.e. Kruskal-Wallis test due to violations of the assumption of normality and homogeneity of variance; ^ higher score indicates worse performance; $p<0.05$=*; ~ sig following corrections for multiple comparisons; PAL = Paired Associate Learning subtest from the CANTAB battery; HVLT-R DI =Hopkins Verbal Learning Test Revised discrimination index; COWAT = controlled oral word association test; TMTB=Trail Making Test Part B; GNT=Graded Naming Test; FFNT=Famous Faces Naming Test; ACE=Addenbrookes Cognitive Examination

Following adjustment of the alpha level in accordance with Holms' sequential Bonferroni correction method (Holm 1979) as a means of correcting for multiple comparisons, significant differences in the baseline performances of aMCI converters and aMCI non-converters were found on the ACE ($t(42) = 2.98$, $p<0.01$, $r= 0.42$) and HVLT-DI ($t(41) = 2.81$, $p <0.01$, $r=0.40$) only.

Only one aMCI participant obtained a score below the 5th%ile (i.e. more than 2 SD below) of our healthy elderly control group at baseline on the GNT. None of the aMCI patients were performing below the 5th%ile of our control group on the GFNT at baseline. Sensitivity, specificity, NPV and PPV values relating to the prediction of conversion from aMCI to dementia were therefore based on a performance of more than 1.5 SD below age norms and can be summarised, in that order, as follows; GNT 0.38, 0.68, 0.43, 0.63; GFNT 0.44, 0.68, 0.5, 0.63. Using a combination of ACE < 88/100 and/or PAL > 14 errors, the overall rate of classification accuracy was 68% with a sensitivity of 72%; specificity of 65%; PPV of 59% and NPV of 77%.

Backward logistic regression analysis with age, NART FSIQ, years of follow-up, and the neuropsychological measures for which baseline performance differentiated converters and non-converters (HVLT-DI and ACE total score), resulted in a final model, completed after 5 iterations, comprising the variables ACE total score and HVLT-DI score only, yielding an overall classification accuracy (aMCI converter vs. aMCI non-converter) of 74%, sensitivity 65%, specificity 80%, NPV 77%, PPV 69%.

Table 6.4 Summary of all variables within the regression equation (Step 1)

	B(SE)	Lower 95th%ile CI Exp(B)	Exp(B)	Upper 95th %ile CI Exp(B)	df	p-value sig
Age at study entry	0.07(0.07)	0.94	1.07	1.23	1	0.32
Length of follow-up	0.04(0.04)	0.97	1.04	1.11	1	0.31
NART FSIQ	0.02(0.06)	0.92	1.02	1.14	1	0.73
ACE total score	-0.17(0.09)	0.71	0.84	1.00	1	0.06
HVLT-R DI	-0.40(0.2)	0.47	0.68	0.98	1	0.04
Constant	8.68(9.66)	-	5903.63	-	1	0.37

Table 6.5 Classification tables for the initial and final steps of the regression analysis

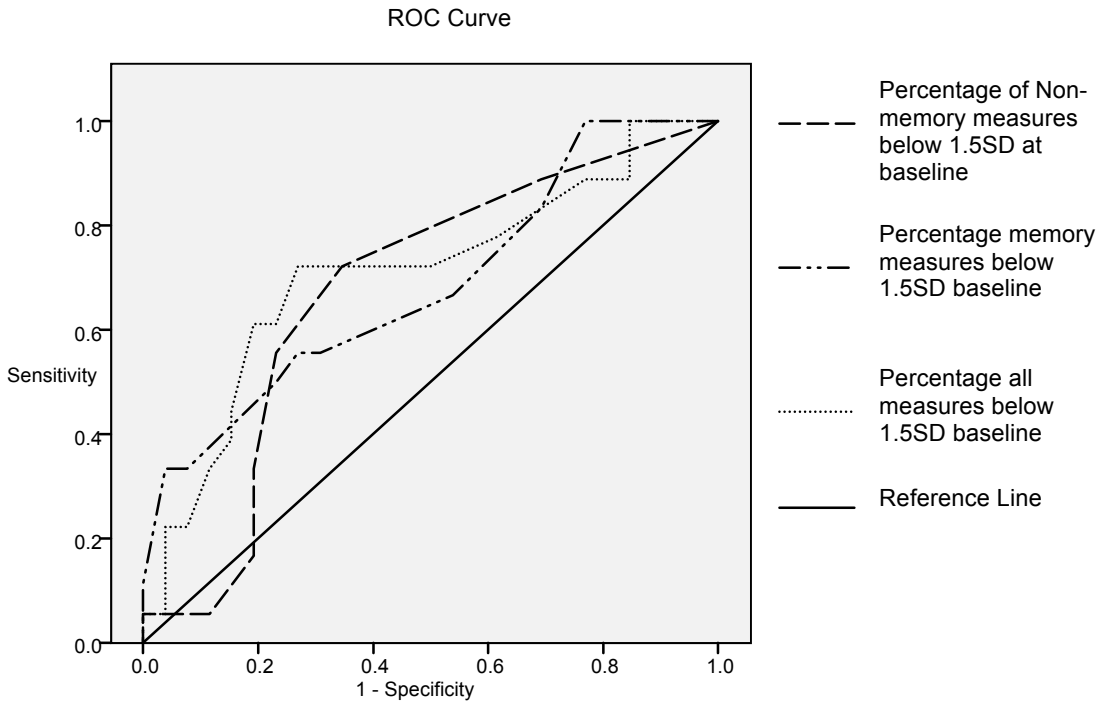
		Predicted		
		Diagnosis		
Observed		No Dementia	Dementia	Percentage Correct
Step 1 Diagnosis	No Dementia	21	4	84.0
	Dementia	7	10	58.8
Overall Percentage				73.8
Step 4 Diagnosis	No Dementia	20	5	80.0
	Dementia	6	11	64.7
Overall Percentage				73.8

Table 6.6 Summary of final regression model

	95% CI for exp b			
	B(SE)	Lower	exp b	Upper
Included				
Constant	15.32* (6.74)		4501208	
ACE Total Score	-0.15* (0.74)	0.75	0.86	1.00
HVLT-R DI	-0.32* (0.16)	0.53	0.73	1.00

R squared = 0.27 (Hosmer & Lemeshow), 0.24 (Cox & Snell), 0.32 (Nagelkerke). Model chi-square(2) = 11.45, $p<0.05$ *

Figure 6.3 Comparative accuracies of endpoint classification (dementia vs. no dementia) for three composite scores



AUC percentage non-memory below 1.5 SD at baseline = 0.68; AUC percentage memory below 1.5 SD at baseline = 0.69; AUC percentage all below 1.5 SD at baseline = 0.71

The percentage of neuropsychological measures on which baseline performance fell 1.5 SD or more below a healthy age and IQ matched control group, was associated with the highest level of overall endpoint classification accuracies (AUC = 0.71). Corresponding AUC's were slightly lower, although similar, for the corresponding non-memory and memory composite scores (AUC = 0.68 & 0.69, respectively).

The high level of agreement (i.e. 87%) observed between the two methods (i.e. 2 or more memory measures 1 SD or more below controls, or 1 or more memory measure 1.5 SD or more below controls) of defining memory impairment in aMCI participants at baseline prevented comparison of conversion rates to dementia across these.

6.1.5 Discussion

Forty one percent of patients who met criteria for aMCI at study entry point received a clinical diagnosis of dementia within the following 4 years, yielding an annual conversion rate of 10% per annum. In light of evidence to suggest that the evolution from mild cognitive impairment to dementia is time-dependent (Busse et al. 2006) comparison in annual conversion rates reported by other studies was restricted to those studies with follow-up periods extending beyond 2.5 years (i.e. 2.5 – 3.5 years). The mean annual conversion rate to dementia among 5 such studies is almost identical (9.7 % per annum) to that observed in the present study.

Performances at baseline on the ACE and HVLT-R DI were able to discriminate between future aMCI converters and non-converters at a time when general levels of cognitive functioning falling above higher level cut off points for dementia (i.e. >27/30 on the MMSE and > 88/100 on the ACE), and classify MCI patient in accordance with their prognostic fate with a moderate degree of overall accuracy. Differences in the baseline performances of the aMCI converter and non-converter groups on these measures were not accountable for in terms of age, FSIQ or years of follow-up, which did not vary significantly between the groups. The average score of the converter group on the HVLT-R DI at baseline equated to a performance 1.75 SD below the published age and education matched control mean (Benedict et al. 1998), and 1.5 SD below the mean of our own healthy elderly age and IQ matched control sample. The corresponding values were 0.58 SD and 0.37 SD for the non-converter group (i.e. both within 1 SD of their respective control means) implying that there is a greater risk of conversion to dementia among the subset of aMCI patients who perform abnormally on this task, and that such patients are readily identifiable on the basis of published norms.

Similar values derived for the ACE, in relation to both published normative (Mathuranath et al. 2000), mean converter = -1.75 SD; mean non-converter = - 0.5 SD, and study control data, mean converter = -2.5 SD; mean non-converter = -1 SD, provide further support for the designation of 88/100 as a higher cut of point for dementia and suggest that use of this cut off score is appropriate among aMCI sufferers, despite the younger age group of the original ACE normative sample.

As such, in clinical practice, the combined performances of aMCI patients on the ACE and HVLT-R DI could be used to inform decisions about the frequency of future contact / monitoring required, or in combination with additional clinical information (i.e. levels of carer rated depressive symptoms (Lu et al. 2009), APOE 4 carrier status (Petersen et al. 2005), corroborative history, neuroimaging findings, family history and/ or qualitative aspects of a patients presentation) to inform whether or not to treat pharmacologically. The relatively small proportion of aMCI patients showing resolution of their cognitive symptoms over time also has implications for the clinical management of such patients, as a number of empirically validated rehabilitative methods for the cognitive rehabilitation of early stage AD are known to exist (Clare and Woods 2004) and could theoretically be used to enhance the day to day memory functioning of patients with aMCI.

Furthermore, although it was not possible to compare the predictive validities of the two methods of determining memory impairment in aMCI, the high levels of agreement between the methods suggests that it is unlikely that a patient classified as aMCI on grounds of a performance of 1 SD or more below the control mean on two memory measures would be re-classified as normal if the criteria of <1.5 SD on one memory measure were applied, and vice versa. As such, the application of either of these methods to define memory impairment in aMCI, within a tertiary specialist memory clinic setting is supported. Similar high levels of agreement in classification as MCI or normal using these two methods of psychometric definition were reported in a recent study by Jak et al. (Jak et al. 2009). The poor to fair AUCs associated with the numbers of memory, non-memory and combined neuropsychological baseline measures performed more than 1.5 SD below age norms, on the other hand, suggests that the use of composite scores (defined in this manner) are of limited clinically utility in the prediction of outcome in aMCI.

Baseline scores on both the HVLT-R DI and the total score on the ACE were significant independent predictors of future conversion to dementia among aMCI sufferers. Closer inspection of the regression analysis reveals that the HVLT-R DI score contributes to the overall classification accuracy of the ACE by increasing negative predictive value i.e. by reducing the probability of conversion to dementia in the face of a sound score. This implies that an impairment of a consolidation/storage nature is generally present in cases where a diagnosis of dementia (AD, VD or AD/VD) is reached within the following 4 years.

The PAL subtest from the CANTAB is also a measure of cued recognition ability. Whilst performance at baseline is reportedly unimpaired in some Questionable Dementia and aMCI clinic based samples (Fowler et al. 1995), and significant differences between future converters have not been found consistently (Ahmed et al. 2008b), a decline across time in the years prior to diagnosis has been established (Fowler et al. 1997). It is possible that the emergence of cued recall impairment is time bound, arising closer to the point at which AD can be diagnosed on clinical grounds and sometime after initial difficulties with the less supported function of free recall become apparent. The ability of cueing to facilitate episodic recall may fade with pre-clinical disease progression, giving rise to an encoding/consolidation profile of memory impairment. In the case of the HVLT-R DI, the inability of aMCI patients to implicitly make use of cues to the same extent as 'healthy' elderly may be amplified by the semantic grouping structure of this measure. Savage et al. have also demonstrated an encoding/consolidation deficit, that is prototypical of AD, in small groups of aMCI and early AD patients using a verbal list learning task (i.e. The California Verbal Learning Test) with similar scope for implicit semantic grouping of list items (Pike et al. 2008).

This observation has implications for the recently proposed new research criteria for AD (Dubois et al. 2007), in which the requirement for objective evidence of significantly impaired episodic memory has been elaborated upon. The new criteria emphasise the superior discriminative and predictive power of delayed as compared to immediate recall measures as well as the importance of establishing an encoding and storage deficit using test paradigms that provide encoding specificity, on grounds that reduced benefit from cueing at recall reliably identifies prodromal AD. Our findings lend support to the specification of episodic memory impairment in this manner. However, the limited range of scores attainable using the HVLT-R DI and resultant potential for floor effects, suggests it may be lesser suited, than PAL, for monitoring decline in episodic memory function over time.

The newer version of the ACE-R (Mioshi et al. 2006) incorporates a delayed cued verbal recognition element. In light of the added predictive value of the HVLT-R DI demonstrated in the present study, it would seem prudent to evaluate whether or not this measure retains its prognostic contribution alongside the ACE-R.

The mean total score on the ACE for those aMCI within our sample who had received a clinical diagnosis of dementia by the end of the study fell just below the level cut off point for dementia (i.e. 87/100). Despite this, for 26% of these patients, baseline ACE scores fell above the higher cut off point for dementia (i.e. 88/100). This would suggest that where the ACE is used as the sole means of determining the likelihood of developing dementia over the proceeding 4 years, up to one quarter of all aMCI patients receive false reassurance of normality. The implications of using the ACE as sole means to determine the presence or absence of clinically significant levels of cognitive impairment are even greater, with the present findings indicating that 62% of patients who fulfil criteria for aMCI obtain scores of 88 or above on the ACE.

We were unable to replicate the high levels of sensitivity, specificity, positive and negative predictive values that have been previously reported in association with combined PAL and ACE scores, the GNT and the GFT (Ahmed et al. 2008a;Blackwell et al. 2004;Thompson et al. 2002). In the case of the latter naming measures, it should be noted that a cut off point of a performance more than 1.5 SD was employed in place of the previously reported 2 SD cut off (Thompson et al. 2002) on account of the fact that only one aMCI participant was performing more than 2 SD below our control mean. This finding likely reflects the shorter follow-up period (i.e. 13.7 months from BL until study endpoint) used in the latter study, resulting in shorter intervals until diagnosis, which are in turn commensurate with the greater magnitudes of impairment observed on these naming tasks. It is plausible that the predictive validity of neuropsychological measures, or their combinations, varies as a function of the number of years prior to diagnosis. This possibility underscores the need for careful consideration of both the length of follow-up and the levels of cognitive functioning of aMCI cohorts at baseline in study comparisons.

There are a number of study limitations worthy of note. Firstly, although a mean follow-up period of over 4 years compares well with previous clinic based studies of longitudinal outcome in aMCI, it remains possible that additional aMCI participants will go on to receive a clinical diagnosis over the longer term. Secondly, the high average pre-morbid IQ and select nature (i.e. tertiary referral, amnestic single and multi-domain & primarily AD endpoint diagnosis) of our aMCI cohort limits generalisation of the study findings beyond groups that are characterised similarly. Although the overwhelming predominant diagnosis of dementia, where reached was of AD or mixed AD/VD type, a small portion of aMCI

participants were diagnosed with vascular dementia. In the case of 17 aMCI participants no neuroimaging was undertaken to exclude a contribution of brain changes of a non-atrophic, i.e. structural or vascular nature to the cognitive complaint. The resultant inclusion of endpoint clinical diagnosis other than that of pure AD, may have influenced the predictive validity of the neuropsychological measures within our battery, although it could be argued, in a more practical sense, that exclusion of aMCI participants on grounds of multiple risk factors is not reflective of clinical reality. Finally, the diagnosis of AD, particularly in its very early stages, is notoriously difficult to make and a definite diagnosis can only be established by way of brain autopsy at post mortem. Furthermore, there is variability in the point at which clinicians arrive at a diagnosis of dementia, despite a common bias towards avoiding false positive diagnoses. As such, it remains possible that for some aMCI patients clinical diagnostic status at the study endpoint was in part reliant on the idiosyncrasies of their attending consultant. Ideally, consensus diagnosis among a number of different specialists, together with longer-term follow-up to achieve autopsy confirmed diagnoses, may have helped to reduce the likelihood of false or unreliable diagnoses, however adequate staffing of the nature required to do this was not available in the current study. In view of this, potential limitations in both the validity and reliability of dementia diagnoses reached by my consultant psychiatry and geriatrician colleagues must be recognised.

6.1.6 Conclusion

Just under half of the participants who met criteria for aMCI at their baseline study assessment received a clinical diagnosis of dementia at some point over the four years that followed. The 10% annual conversion rate to dementia observed in the present study is similar to rates reported previously in association with clinical based aMCI samples across 2-5-3.5 year follow-up intervals. The vast majority of aMCI patients within our sample displayed persisting or progressive impairment of a cognitive nature over the course of the study. Performance on two neuropsychological measures at baseline differentiated between future aMCI converters and non-converters and performance was predictive of outcome (i.e. dementia vs. no dementia) over and above demographic indices, with an overall classification accuracy of 74%. Over one quarter of aMCI participants who went on to receive a clinical diagnosis of dementia performed above higher level cut off points for dementia on the ACE at their initial assessment. The pervasiveness of the memory impairment in aMCI, the high risk of conversion from aMCI to dementia and the predictive validity of several cognitive measures have implications for the clinical management of patients with aMCI.

Table 6.7 Summary of clinic based findings pertaining to the prediction of dementia using neuropsychological measures

Author (Date)	*Predictors*	*FU yrs*	*Mean MMSE/ACE at BL*	*Mean Age*	*n*	*Mean NART FSIQ Mean education level (years)*	*Cut off*	*% total C (% C per annum)*	*Predictive Value(s)*
Thompson (2002)	GFNT	1.1	27.4	66.6	28	117 / 13	-2SD	25 (22)	GFNT PPV=0.6 NPV=0.94 Sens=86 Spec=81 GNT PPV=1.0 NPV=0.78 Sens=0.14 Spec=1.0
Ahmed (2008)	ACE ALB buildings & patterns	1	C=25.7 / 77.3 NC=29 / 86	71	18	C=11.9 NC=14	<88 >14 errors i.e. -2SD	39 (39)	ACE and / or PAL Sens=100% Spec=82% PPV=78% NPV=100%
Lehmer (2005)	MMSE Block span [AKT] Digit symbol Misplaced objects Name face association SR total SR delay	2	C=25.8 NC=28	C=71 NC=66	107	C=10 NC=12	<7	40 (20)	Selective Reminding Test delayed recall Sens >80 Spec >80 AUC = 0.94 PPV < 0.40
Griffith (2006)	DRS total Semantic fluency DRS memory VR II VR % retention DRS init/perseveration	2	28	C=70 NC=67	49	13	<37 <26%	34.22 (17)	Dementia Rating Scale initiation/perseveration score & WMS-III VR % retention Classification accuracy = 85.7% Sens=76.9% Spec=88.9%
Amieva (2004)	Age MMSE total MMSE word recall BVRT IST DSST LCT	2	27	C=73 NC=68	90	89% primary school diploma			LCT only standalone predictor in regression model
Schmidtke (2007)	None	1.6	C=25.7 NC=26.6	C=76 NC=73	75	C=10.2 NC=9.6	NA		NA
Perri (2007)	Word list recall all indices	2	C=26.3 NC=27.7	C=73 NC=68	190	C=7.5 NC=7.7	-1.5 SD	41.5 (20.8)	Cumulative delayed recall index sens 75% spec 68.5%
Tabert (2006)	SRT WMS-VR BVRT recog BNT ANT BD OA Digit symbol CFL	3.5	C=26.1 NC=28.1	C=73 NC=64	115	C=13.9 NC=15		48 (13.7)	Selective Reminding Test total immediate recall score Digit symbol time to completion Sens 76% Spec 90%

	Similarities Mattis identities and oddities								accuracy 86% PPV 76 NPV 90%
Lee (2006)	CERAD word list subtests and constructional recall subtest MMSE CDR	3	C=25 NC=NA	C=71 NC=74	72			19.4 (6)	NA
Loewenstein (2007)	SIT OME	3	26	77	76	C=14 NC=12	<4	35.5 (12)	SIT recall PPV=70.4% NPV=73.5% AUC=77.5
Estevez-Gonzalez (2004)	MMSE Age Face naming	2	C=26.3 NC=27.9	C=73 NC=66	53	C=7.1 NC=8.1	NA	48 (24)	NA
Albert (2001)	CVLT total CSRT TMTB SOT Alpha span FAS	3	29	72	123	14	NA	19 (6.3)	TMTB, WMS-R VR immediate recall figures, SOT total score overall accuracy= 80% Sens = .074 Spec = 0.83
Fox (1998) Pre-clinical familiar AD	Performance IQ RMT words	6	C=29 NC=29	44	63	100	NA	16	NA
Blackwell (2004)	NART ADAS-cog MMSE RMT words RMT faces CANTAB Pattern recognition Doors recognition CANTAB delayed matching to sample CANTAB PAL WMS-R LMII GNT New semantic naming battery Category fluency	2.5	C=25 NC=29	C=72 NC=62	43	C=117 NC=119	NA	26 (10.4)	PAL and age PPV= 81% NPV=96.6% PAL, age and GNT overall accuracy=100%
Estevez-Gonzalez (2003)	All indices of the RAVLT except learning	2	C=26.3 NC=27.9	C=73 NC=66	70	C=7.1 NC=8.1	NA		NA
Rami (2007)	aMCI +prAD (age, visual memory) aMCI only (delayed memory test, animal fluency)	1	26 aMCI 24 prAD	73	48	7-8	NA	20 (20)	Logistic reg aMCI + prAD visual memory and age were sig predictors Logistic reg aMCI alone no sig predictors of C No details of model or accuracy of prediction provided

ACE, Addenbrookes Cognitive Examination; ADAS, Alzheimer's Disease Assessment Scale; AKT, The Alters-Konzentrations Test; ALB, Associative learning battery; ANT, Animal Naming Test; BD, Block Design; BNT, Boston Naming Test; BVRT, Benton Visual Retention Test; CANTAB, Cambridge Automated Neuropsychological testing Assessment Battery; CERAD, Consortium to Establish a Registry for Alzheimer's Disease; CFL, lexical fluency using these letters; C, converter; CSRT, Cued and Selective Reminding test; CVLT, California Verbal Learning Test; DRS, Dementia Rating Scale; DSST, Digit Symbol Substitution Test; FAS, lexical fluency using these letters; FU, follow-up; GNT, Graded Naming Test; GFNT, Graded Faces Naming Test; IST, Isaacs Set Test; LCT, Letter Cancellation Test; MMSE, Mini Mental State Examination; n, sample size; NC, non-converter; OA, Object Assembly; OME, Object Memory Evaluation; RAVLT, Rey Auditory Verbal Learning Test; RMT, Recognition Memory Test; SIT, Semantic Interference Test; SOT, Self Ordering Test; SRT, Selective Reminding Test; TMTB, Trail Making Test Part B; VR, Visual Reproduction; WMS, Wechsler Memory Scale; reg, regression; AUC, area under the curve; NPV, negative predictive value; PPV, positive predictive value; Spec, specificity; Sens, sensitivity; Conv, converter; Non-conv, non-converter; prAD, prodromal Alzheimer's Disease.

7. General Summary and Conclusions

Elderly patients with memory complaints, who exhibit cognitive impairment(s) that fall short of dementia, typically display deficits of verbal episodic and semantic memory on comprehensive neuropsychological assessment. Many such patients perform above higher-level cut-off points for dementia on even the most comprehensive bedside cognitive screening measures at initial presentation. Patients presenting in this manner, comprise around one quarter of the referrals received by tertiary referral memory clinics. Hence, there is a clear need for evidence based guidance as to how they might best be managed clinically, with regards to assessment, investigative and treatment requirements.

The findings of the this series of studies advance our current understanding of the neuropsychology of early and preclinical AD and provide an evidence base from which clinical management decisions pertaining to aMCI could be drawn.

With regards to the assessment and detection of aMCI, both the MMSE and the ACE lack sensitivity to the cognitive impairment that underlies aMCI. This implies that large numbers of elderly patients presenting with memory complaints who are cognitively evaluated with these bedside cognitive screening instruments, may be mistakenly categorised as 'worried well' or 'normal for age'. Our findings highlight the need for comprehensive neuropsychological assessment, beyond bedside cognitive screening, that incorporates a number of measures of functioning within the domains of episodic, semantic memory and executive functioning, to obtain acceptably high levels of sensitivity to aMCI. The sensitivity of a number of such measures to aMCI is well-documented in this, as well as other studies (Ahmed et al. 2008a;Ahmed et al. 2008b;Alladi et al. 2006;Dudas et al. 2005b). Measures of face and object naming, category fluency and speeded divided attention, in conjunction with those assessing of delayed verbal and visual recall and recognition memory are among the most sensitive, although none of these retains acceptable levels of specificity in differentiating aMCI from normal aging or depression. These measures can, however, be applied to the differential diagnosis of AD and normal aging with very high levels of accuracy.

The study findings have implications for the provision of information of a prognostic nature to patients with aMCI. In just under 50% of cases where criteria for aMCI were met at the study baseline assessment, a diagnosis of dementia was reached on clinical grounds within the following four years. There would appear to be some consistency in the annual conversion rates from aMCI to dementia reported across a range of clinic based longitudinal studies of similar length, despite some minor differences in the manner in which aMCI criteria are applied. Equally, following exclusion of 4 aMCI study participants with an unstable neuropsychological profile (i.e. impaired at the point of recruitment but not baseline study assessment), the rate of reversion to 'normal' during the course of the study (defined psychometrically in accordance with chance levels of poor test scores) appears to be low (15%). For the vast majority of aMCI who do not receive a clinical diagnosis of dementia cognitive impairment either persists or worsens over time.

On the basis of our own findings and similar clinic based studies, it would therefore seem reasonable to assign a one in two chance of developing dementia across the proceeding 4 year period to patients who have been identified within specialist clinics as fulfilling criteria for aMCI on more than one occasion. Where conversion to dementia does not occur, one could expect to see persisting or worsening cognition in a majority of cases.

In the present study, along with others of a similar duration, a high rate of conversion from aMCI to dementia was observed. The predominant alternative outcome to dementia in the present study was persisting or worsening cognitive function. Two easily administered cognitive tasks contributed significantly to the identification of aMCI as it represents pre-clinical dementia. Together, these findings challenge the current standard clinical practice wherein aMCI patients are not routinely placed under review, or referred on for further investigations, and are provided with little, if anything, in the way of intervention(s) or prognostic information.

The present study findings indicate, along with that of others, (Fowler et al. 1995;Fowler et al. 1997;Fowler et al. 2002) that it is possible to detect cognitive decline of an 'abnormal' magnitude on measures of neuropsychological function with relevant knowledge of re-test reliabilities, effects of normal aging and practice effects. In view of this, and the high likelihood that clinically significant levels of progressive cognitive deterioration will be

observed in aMCI over time, further consideration should be given to the manner and frequency with which aMCI patients are reviewed in clinical practice.

At the very least, the initial clinical assessment of aMCI could include measures such as the ACE and HVLT-R DI with established prognostic utility. Where indicated and available, the high risk aMCI patients could receive further investigations in the form of imaging (i.e. PET) and or genetic testing, where these are available. Outcomes from these investigations could be used together with information of a qualitative/clinical nature to further refine ascertainment of prognosis and/or the appropriate timing of pharmacological intervention.

Finally, the high rates of conversion from aMCI to dementia, and of persisting and progressive cognitive decline in aMCI, suggest that there may be a role for interventions of a non-pharmacological nature in aMCI. There exist a number of empirically validated internal and external strategies to aid everyday memory function in early AD that could theoretically be extended to aMCI (Clare and Woods 2004). Furthermore, it is conceivable that there may be advantages in introducing such strategies at the earliest possible stage of the disease process in AD, at a time when some capacity to learn remains present, affording the opportunity to establish new routine(s) prior to the onset of more generalised impairment of cognition and the emergence of co-morbid behavioural symptoms. The introduction of interventions of this nature at an earlier stage would also maximise the patient's capacity to contribute to and shape these in accordance with their own personal wishes.

Prior to concluding, there are a number of study strengths and limitations worthy of note. Perhaps the most notable study strengths relate to the large number of participants comprising the aMCI group and the relatively long follow-up period (in comparison to other published clinic-based neuropsychological longitudinal research). The latter serving to reduce the likelihood of false negative endpoint classifications of non-converter among aMCI patients who do in fact go on to develop dementia post study endpoint. The study represents one of, if not *the*, most detailed and comprehensive longitudinal neuropsychological evaluation of cognition in aMCI. The study battery taps different aspects of functioning within specific cognitive domains (i.e. visual and verbal memory, free and cued delayed recall) allowing for a more detailed comparison of the prognostic and differential diagnostic performance of a number of different neuropsychological measures

that are commonly used in the assessment of the elderly. For many such measures, the study represents a first attempt to compare their differential diagnostic utility and prognostic validity in the assessment of aMCI, early AD and depression. The heterogeneity of aMCI is widely acknowledged. In the present study, however, the tertiary nature of referrals mean't that all aMCI participants had undergone psychiatric/ or physician and GP examinations prior to study recruitment with the effect of minimising the likelihood that cognitive complaints were related to underlying medical illness or were sufficient to fulfil criteria for early dementia. The minimal impairment of general levels of cognitive functioning of our aMCI group further reduced likelihood of observing 'artificial' inflation of the prognostic significance and effect sizes of tests, on account of a proportion of this group already meeting criteria for early AD.

Study limitations worthy of note include the fact that study outcome criteria were incomplete owing to an inability to collect a number of outstanding functional scales for the aMCI folk. Furthermore, the clinical outcome of dementia vs. no dementia was reliant on the judgement of the attending Consultant Psychiatrist and was not independently verified via application of NINCDS or DSM-IV criteria. As such, and without autopsy confirmation of AD or other forms of dementia, outcome criteria in accordance with the medical file method, most accurately represent the numbers of aMCI patients who received a diagnosis of dementia as opposed to numbers who converted to dementia.

A further criticism of the research involves the reliance on neuropsychological measures to recruit and in the case of early AD and aMCI, define participant groups. There is a degree of circularity in using cognitive complaints and neuropsychological tests to define clinical conditions and then to compare them which must be acknowledged here. The neuropsychological assessments were carried out by a numer of different raters across the course of the study, without information regarding inter-rater reliability. As such, it is conceivable that some of the variability in performances of individual participants at different time points is accountable for in terms of differences in rater test administration styles, techniques or scoring (despite the use of a common administration manual for all of the neuropsychological measures was employed as a means of maximising inter-rater reliability). Functional scales were administered to the spouses or primary carers of the aMCI participants at the studyend point only. This prevented the tracking of functional status across time. Consequently, functional decline had to be infered from the absolute

scores on the scale at study endpoint, as opposed to change scores from baseline to final assessment. Furthermore, the Geriatric Depression Scale was administered to the depression group only, such that it is conceivable that elevated levels of depressive symptoms were also present in the aMCI, AD and or CT groups with the effect of complicating the interpretation of the relative performances of the groups. With further regards to the depression group, limited clinical details were gathered. This lack of information prevented fuller interpretation of neuropsychological performance in accordance with whether predominant make up is EOD or LOD, a distinction that is known to influence the neuropsychological performance of elderly patients with depression.

This study highlights the prognostic value of a measure of delayed verbal recognition in the prediction of conversion from aMCI to dementia. This measure, together with the original version of the ACE, added over and above the clinical variables of age, FSIQ and duration of follow-up, to the prediction of dementia. As the revised version of the ACE (ACE-R) contains a verbal recognition subtest, it would seem worthwhile determining whether the inclusion of such a subtest negated the additional predictive power of the discrimination index score from the HVLT-R.

Furthermore, we demonstrated that verbal fluency measures, and more specifically the differential performance in lexical and semantic measures, better discriminate between the healthy elderly and those with aMC, than traditional measures of delayed verbal recall. The ACE also demonstrated greater (although still inadequate) sensitivity to aMCI than the MMSE. Although classified as a bedside cognitive screening measure, the ACE is relatively lengthy compared to the MMSE, requiring on average 15 mintues to administer. This is a limiting factor in its use in busy clinic settings. As such, it would be of practical import to determine the extent to which its discriminative capacity and sensitivity to aMCI is attributable to performance on the fluency subtests alone (which require only 2 mintues to administer).

The present study focuses soley on performance on neuropsychological measures as a means of detecting incipent dementia. As dissuced previously however, there are a number of different lines of investigation comprising behavioural, imaging and CSF biomarkers for pre-clinical AD. Future research should pursue a more integrated multi-disciplinary approach,

allowing for the determination of the prognostic power of combinations of all of the above markers and direct comparison of their respective prognostic values in identifying pre-clinical AD.

Clinicians views regarding the diagnostic value of aMCI and the role of neuropsychology vs. clinical opinion in arriving at such a 'diagnosis', remain divided. On the one hand, clinical acumen is difficult to teach, arising largely out of experience. Clinical judgement is equally difficult to measure, and methods of arriving at a clinical judgement are likely to vary widely from one practitioner to the next, making it difficult to research and compare findings across different studies. On the other hand, neuropsychological expertise, and indeed the time to conduct additional cognitive measures within the course of a medical, psychiatric or neurological examination, is not always available. Furthermore, performance on neuropsychological testing may be influenced by a number of factors (i.e. motivation, fatigue, anxiety, sensory deficits ect.) that do not relate to the pathology under investigation. This together with the scarcity of comprehensive local normative data for a number of neuropsychological measures may complicate interpretation of test scores.

Regardless of the means by which a diagnosis of aMCI is arrived at, the cognitive features of a patient's neuropsychological presentation remain vital components of the very early and differential diagnostic process. The above studies demonstrate that there remains considerable scope to improve upon the manner in which cognitive aspects of AD and aMCI criteria are implemented, in order to facilitate earlier diagnosis and optimise the clinical management of these patient groups.

8. References

Reference List

1. 2002. Bibliography - Current World Literature. *Current Opinion in Neurology* **15**: 499-518.

2. Adlam ALR, Bozeat S, Arnold R, Watson P, Hodges JR. 2006. Semantic knowledge in mild cognitive impairment and mild Alzheimer's disease. *Cortex* **42**(5): 675-684.

3. Aggarwal NT, Wilson RS, Beck TL, Bienias JL, Bennett DA. 2005. Mild cognitive impairment in different functional domains and incident Alzheimer's disease. *J Neurol Neurosurg Psychiatry* **76**(11): 1479-1484.

4. Aharonson V, Halperin I, Korczyn AD. 2007. Computerized diagnosis of mild cognitive impairment. *Alzheimer's and Dementia* **3**(1): 23-27.

5. Ahmed S, Arnold R, Thompson SA, Graham KS, Hodges JR. 2008a. Naming of objects, faces and buildings in mild cognitive impairment. *Cortex* : 746-752.

6. Ahmed S, Mitchell J, Arnold R, Nestor PJ, Hodges JR. 2008b. Predicting rapid clinical progression in amnestic Mild Cognitive Impairment. *Dementia & Geriatric Cognitive Disorders* **25**: 170-177.

7. Akhtar S, Moulin CJA, Bowie PCW. 2006. Are people with mild cognitive impairment aware of the benefits of errorless learning? *Neuropsychological Rehabilitation* **16**(3): 329-346.

8. Albert M, Blacker D, Moss MB, Tanzi R. 2007. Longitudinal change in cognitive performance among individuals with mild cognitive impairment. *Neuropsychology* **21**(2): 158-169.

9. Albert MS, Moss MB, Tanzi R, Jones K. 2001. Preclinical prediction of AD using neuropsychological tests. *Journal of the International Neuropsychological Society* **7**(5): 631-639.

10. Alladi S, Arnold R, Mitchell J, Nestor PJ, Hodges JR. 2006. Mild cognitive impairment: applicability of research criteria in a memory clinic and characterization of cognitive profile. *Psychol Med* **36**(4): 507-513.

11. American Psychiatric Association. 1994. *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association: Washington, DC.

12. Amieva H, Jacqmin-Gadda H, Orgogozo J, Le Carret N, Helmer C, Letenneur L, Barberger-Gateau P, Fabrigoule C, Dartigues J. 2005. The 9 year cognitive decline before dementia of the Alzheimer type: a prospective population-based study. *Brain* **128**: 1093-1101.

13. Amieva H, Letenneur L, Dartigues JL, Rouch-Leroyer I, Sourgen C, Alchee-Biree F. 2004. Annual rate and predictors of conversion to dementia in subjects

presenting mild cognitive impairment criteria defined according to a population-based study. *Dement Geriatr Cogn Disord* **18**(1): 87-93.

14. Archer HA, MacFarlane F, Price S, Moore EK, Pepple T, Cutler D, Frost C, Fox NC, Rossor MN. 2006. Do symptoms of memory impairment correspond to cognitive impairment: a cross sectional study of a clinical cohort. *International Journal of Geriatric Psychiatry* **21**: 1206-1212.

15. Arnaiz E, Blomberg M, Fernaeus S-E, Wahlund L-O, Winbald B, Almkvist O. 2000. Psychometric Discrimination of Alzheimer's Disease and Mild Cognitive Impairment. *Alzheimer's Reports* **3**(2): 97-103.

16. Arnaiz E, Almkvist O. 2003. Neuropsychological features of mild cognitive impairment and preclinical Alzheimer's disease. *Acta Neurol Scand* **107**(179): 34-41.

17. Arnaiz E, Almkvist O, Ivnik RJ, Tangalos EG, Wahlund LO, Winblad B, Petersen RC. 2004. Mild cognitive impairment: a cross-national comparison. *J Neurol Neurosurg Psychiatry* **75**(9): 1275-1280.

18. Arnett PA, Higginson CI, Voss WD, Bender WI, Wurst JM, Tippin JM. 1999. Depression in multiple sclerosis: relationship to working memory capacity. *Neuropsychology* **13**: 546-556.

19. Artero S, Petersen R, Touchon J, Ritchie K. 2006. Revised criteria for mild cognitive impairment: validation within a longitudinal population study. *Dement Geriatr Cogn Disord* **22**: 465-470.

20. Artero S, Tierney MC, Touchon J, Ritchie K. 2003. Prediction of transition from cognitive impairment to senile dementia: A prospective longitudinal study. *Acta Psychiatrica Scandinavia* **107**(5): 390-393.

21. Attix DK, Welsh-Bohmer KA. 2006. *Geriatric Neuropsychology*. The Guilford Press: New York.

22. Auriacombe S, Lechevallier N, Amieva H, Harston S, Raoux N, Dartigues JF. 2006. A Longitudinal Study of Quantitative and Qualitative Features of Category Verbal Fluency in Incident Alzheimer's Disease Subjects: Results from the PAQUID Study. *Dement Geriatr Cogn Disord* **21**(4): 260-266.

23. Backman L, Jones S, Berger A, Small BJ. 2005. Cognitive impairment in prelcinical alzheimer's disease: a meta-analysis. *Neuropsychology* **19**(4): 520-531.

24. Baddeley A, Della Sala S, Papagno C, Spinnler H. 1997. Testing Central Executive Functioning with a Pencil and paper Test. In *Methodology of frontal and executive function*, Rabbitt PMA (ed). Psychology Press: 61.

25. Baddeley A, Hitch GJ. 1974. Working memory. In *Recent advances in learning and motivation*, Bower G (ed). Academic Press: London; 40-79.

26. Baddeley AD, Baddeley HA, Bucks RS, Wilcock GK. 2001. Attentional control in Alzheimer's disease. *Brain* **124**(8): 1492-1508.

27. Baddeley AD, Bressi S, Della Sala S, Logie R, Spinnler H. 1991. The decline of working memory in Alzheimer's disease: a longtitudinal study. *Brain* **114**: 2521-2542.

28. Baddeley AD, Logie R, Brssi S, Della Sala S, Spinnler H. 1986. Dementia and working memory. *Quarterly Journal of Experimental Psychology* **38A**: 603-618.

29. Bak TH, Rogers TT, Crawford LM, Hearn VC, Mathuranath PS, Hodges JR. 2005. Cognitive bedside assessment in atypical parkinsonian syndromes. *J Neurol Neurosurg Psychiatry* **76**: 420-422.

30. Balthazar MLF, Martinelli JE, Cendes F, Damasceno BP. 2007. Lexical semantic memory in amnestic mild cognitive impairment and mild Alzheimer's Disease. *Arquivos de Neuro-Psiquiatria* **65**(3A): 619-622.

31. Barbeau E, Didic M, Tramoni E, Felician O, Joubert S, Sontheimer A, Ceccaldi M, Poncet M. 2004. Evaluation of visual recognition memory in MCI patients. *Neurology* **62**(8): 1317-1322.

32. Bartolini M, Coccia M, Luzzi S, Provinciali L, Ceravolo MG. 2005. Motivational symptoms of depression mask preclinical Alzheimer's disease in elderly subjects. *Dementia and Geriatric Disorders* **19**: 31-36.

33. Basso MR, Bronstein RA, Lang JM. 1999. Practice effects on commonly used measures of executive function across twelve months. *The Clinical Neuropsychologist* **13**: 283-292.

34. Baudic S, Barba GD, Thibaudet MC, Smagghe A, Remy P, Traykov L. 2006. Executive function deficits in early Alzheimer's disease and their relations with episodic memory. *Archives of Clinical Neuropsychology* **21**(1): 15-21.

35. Beatty WW, Salmon DP, Troester AI, Tivis RD. 2002. Do primary and supplementary measures of semantic memory predict cognitive decline by patients with Alzheimer's Disease? *Aging Neuropsychology And Cognition* **9**(1): 1-10.

36. Belleville S, Chertkow H, Gauthier S. 2007. Working memory and control of attention in persons with Alzheimer's disease and Mild Cognitive Impairment. *Neuropsychology* **21**(4): 458-469.

37. Benedict RHB, Schretlen D, Groninger L, Brandt J. 1998. Hopkins Verbal Learning Test - Revised: Normative data and analysis of inter-form and test-retest reliability. *The Clinical Neuropsychologist* **12**(1): 43-55.

38. Bennett DA, Wilson RS, Schneider JA, Evans DA, Beckett LA, Aggarwal NT, Barnes LL, Fox JH, Bach J. 2002. Natural history of mild cognitive impairment in older persons. *Neurology* **59**(2): 198-205.

39. Bennett IJ, Golob EJ, Parker ES, Starr A. 2006. Memory evaluation in mild cognitive impairment using recall and recognition tests. *Journal of Clinical & Experimental Neuropsychology* **28**(8): 1408-1422.

40. Berry DT, Allen RS, Schmitt FA. 1991. The Rey-Osterrieth Complex Figure: Psychometric characteristics in a geriatric sample. *The Clinical Neuropsychologist* **5**: 143-153.

41. Bier JC, Donckels V, Van Eyll E, Claes T, Slama H, Fery P, Vokaer M. 2005. The French Addenbrooke's cognitive examination is effective in detecting dementia in a French-speaking population. *Dement Geriatr Cogn Disord* **19**(1): 15-17.

42. Bier JC, Ventura M, Donckels V, Eyll E, Claes T, Slama H, Fery P, Vokaer M, Pandolfo M. 2004. Is the Addenbrookes Cognitive Examination effective to detect frontotemporal dementia? *J Neurol* **251**(4): 428-431.

43. Bird CM, Cipolotti L. 2007. The utility of the recognition memory test and the graded naming test for monitoring neurological patients. *British Journal of Clinical Psychology* **46**: 223-234.

44. Bird CM, Papadopoulou K, Ricciardelli P, Rossor MN, Cipolotti L. 2004. Monitoring cognitive changes: Psychometric properties of six cognitive tests. *British Journal of Clinical Psychology* **43**: 197-210.

45. Black SE. 1996. Focal Cortical Atrophy Syndromes. *Brain and Cognition* **31**(2): 188-229.

46. Blacker D, Lee H, Muzikansky A, Martin EC, Tanzi R, McArdle JJ, Moss M, Albert M. 2007. Neuropsychological Measures in Normal Individuals That Predict Subsequent Cognitive Decline. *Arch Neurol* **64**(6): 862-871.

47. Blackwell AD, Sahakian BJ, Vesey R, Semple JM, Robbins TW, Hodges JR. 2004. Detecting dementia: Novel neuropsychological markers of preclinical Alzheimer's disease. *Dementia & Geriatric Cognitive Disorders* **17**((1-2)): 42-48.

48. Bland JM, Altman DG. 1986. Statistical Methods for assessing agreement between two methods of clinical measurement. *The Lancet* **8**: 307-310.

49. Blossom CMS, Matthews FE, McKeith IG, Bond J, Brayne C. 2007. Early cognitive change in the general population: how do different definitions work? *Journal of the American Geriatric Society* **55**: 1534-1540.

50. Bondi MW, Jak AJ, Delano-Wood L, Jacobson MW, Delis DC, Salmon DP. 2008. Neuropsychological contributions to the early identification of Alzheimer's disease. *Neuropsychology Review* **18**: 73-90.

51. Bozoki A, Giordani B, Heidebrink JL, Berent S, Foster NL. 2001. Mild Cognitive Impairments Predict Dementia in Nondemented Elderly Patients With Memory Loss. *Arch Neurol* **58**(3): 411-416.

52. Braak H, Braak E. 1991. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica* **82**: 239-259.

53. Brandt J. 1991. The Hopkins Verbal Learning Test: Development of a new memory test with 6 equivalent forms. *The Clinical Neuropsychologist* **5**(2): 125-142.

54. Brink TA, Yesavage JA, Lum O, Heersema P, Adey M, Rose TL. 1982. Screening tests for geriatric depression. *Clinical Gerontologist* **1**: 37-44.

55. Brooks BL, Iverson GL, White T. 2007. Substantial risk of "Accidental MCI" in healthy older adults: Base rates of low memory scores in neuropsychological assessment. *Journal of the International Neuropsychological Society* **13**(3): 490-500.

56. Busse A, Bischkopf J, Riedel-Heller SG, Angermeyer MC. 2003. Subclassifications for mild cognitive impairment: prevalence and predictive validity. *Psychological Medicine* **33**(6): 1029-1038.

57. Busse A, Angermeyer MC, Riedel-Heller SG. 2006. Progression of mild cognitive impairment to dementia: a challenge to current thinking. *British Journal of Psychiatry* **189**: 399-404.

58. Caine D, Hodges JR. 2001. Heterogeneity of Semantic and Visuospatial Deficits in Early Alzheimer's Disease. *Neuropsychology* **15**(2): 155-164.

59. Callahan CM, Unverzagt FW, Hui SL, Perkins AJ, Hendrie HC. 2002. Six-Item screener to identify cognitive impairment among potential subjects for clinical research. *Medical Care* **40**(9): 771-781.

60. Cargin JW, Maruff P, Collie A, Masters C. 2006. Mild memory impairment in healthy older adults is distinct from normal aging. *Brain and Cognition* **60**(2): 146-155.

61. Cerhan JH, Ivnik RJ, Smith GE, Tangalos EC, Petersen RC, Boeve BF. 2002. Diagnostic utility of letter fluency, category fluency, and fluency difference scores in Alzheimer's disease. *Clin Neuropsychol* **16**(1): 35-42.

62. Chen P, Ratcliff G, Phil D, Belle SH, Cauley JA, DeKosky ST, Ganguli M. 2000. Cognitive tests that best discriminate between presymptomatic AD and those who remain nondemented. *Neurology* **55**(2 of 2): 1847-1853.

63. Chertkow H, Nasreddine Z, Joanette Y, Drolet V, Kirk J, Massoud F, Belleville S, Bergman H. 2007. Mild cognitive impairment and cognitive impairment, no dementia: Part A, concept and diagnosis. *Alzheimer's and Dementia 3(4)()(pp 266-282), 2007 Date of Publication: Oct 2007* (4): 266-282.

64. Chosak RJ. 2000. Measuring cognitive processes underlying picture naming in Alzheimer's and cerebrovascular dementia: A general processing tree approach. *Journal of Clinical and Experimental Neuropsychology* **22**: 351-369.

65. Christensen H, Griffiths K, Mackinnon A, Jacomb P. 1997. A quantitative review of cognitive deficits in depression and Alzheimer-type dementia. *Journal of the International Neuropsychological Society* **3**: 631-651.

66. Clague F, Dudas RB, Thompson SA, Graham KS, Hodges JR. 2005. Multidimensional measures of person knowledge and spatial associative learning: can these be applied to the differentiation of Alzheimer's disease

from frontotemporal and vascular dementia? *Neuropsychologia* **43**(9): 1338-1350.

67. Clare L, Woods RT. 2004. Cognitive training and cognitive rehabilitation for people with early-stage AD: A review. *Neuropsychological Rehabilitation* **14**(385): 401.

68. Cohen NJ, Eichenbaum H. 1993. *Memory, amnesia, and the hippocampal system*. The MIT Press: Massachusetts, USA.

69. Cohen NJ, Ryan J, Hunt C, Romine L, Wszalek T, Nash C. 1999. Hippocampal systems and declarative (relational) memory: Summarizing the data from functional neuroimaging studies. *Hippocampus* **91**(1): 83-98.

70. Collette F, Linden MVd, Bechet S, Salmon E. 1999. Phonological loop and central executive functioningin Alzheimers disease. *Neuropsychologia* **37**(8): 905-918.

71. Collie A, Maruff P. 2002. An analysis of systems of classifying mild cognitive impairment in older people. *Australian and New Zealand Journal of Psychiatry* **36**(1): 133-140.

72. Collie A, Maruff P, Currie J. 2002. Behavioural Characterization of Mild Cognitive Impairment. *Journal of Clinical and Experimental Neuropsychology* **24**(6): 720-733.

73. Collie A, Maruff P, Shafiq-Antonacci R, Smith M, Hallup M, Schofield PR, Masters CL, Currie J. 2001. Memory decline in healthy older people: Implications for identifying mild cognitive impairment. *Neurology* **56**(1 of 2): 1533-1538.

74. Collie A, Maruff P. 2000. The neuropsychology of preclinical Alzheimer's disease and mild cognitive impairment. *Neuroscience & Biobehavioral Reviews* **24**(3): 365-374.

75. Crawford JR, Garthwaite PH. 2006. Comparing patients' predicted test scores from a regression equation with their obtained scores: A significance test and point estimate of abnormality with accompanying confidence limits. *Neuropsychology* .

76. Crossley M, Hiscock M. 1992. Age-related differences in concurrent-task performance of normal adults: evidence for a decline in processing resources. *Psychology and Aging* **7**: 499-506.

77. Crossley M, Hiscock M, Foreman JB. 2004. Dual-task performance in early stage dementia: differential effects for automatized and effortful processing. *Journal of Clinical & Experimental Neuropsychology* **26**: 332-346.

78. Crowell TA, Luis CA, Vanderploeg RD, Schinka JA, Mullan M. 2002. Memory patterns and executive functioning in mild cognitive impairment and Alzheimer's disease. *Aging Neuropsychology And Cognition* **9**(4): 288-297.

79. Daly E, Zaitchik D, Copeland M, Schmahmann J, Gunther J, Albert M. 2000. Predicting Conversion to Alzheimer Disease Using Standardized Clinical Information. *Arch Neurol* **57**(5): 675-680.

80. Darby D, Maruff P, Collie A, McStephen M. 2002. Mild cognitive impairment can be detected by multiple assessments in a single day. *Neurology* **59**(7): 1042-1046.

81. De Jager C, Blackwell AD, Budge MM, Sahakian BJ. 2005. Predicting cognitive decline in healthy older adults. *Am J Geriatr Psychiatry* **13**(8): 735-740.

82. De Jager CA, Budge MM. 2005. Stability and predictability of the classification of mild cognitive impairment as assessed by episodic memory test performance over time. *Neurocase* **11**(1): 72-79.

83. De Jager CA, Hogervorst E, Combrinck M, Budge MM. 2003. Sensitivity and specificity of neuropsychological tests for mild cognitive impairment, vascular cognitive impairment and Alzheimer's disease. *Psychological Medicine* **33**(6): 1039-1050.

84. de Rotrou J, Wenisch E, Chausson C, Dray F, Faucounau V, Rigaud AS. 2005. Accidental MCI in healthy subjects: a prospective longitudinal study. *European Journal of Neurology* **12**(11): 879-885.

85. DeJager CA, Milwain E, Budge M. 2002. Early detection of isolated memory deficits in the elderly: the need for more sensitive neuropsychological tests. *Psychological Medicine* **32**: 483-491.

86. Della Sala S. 2005. Paper and Pencil Dual Task. In .

87. Della Sala S, Baddeley A, Papagno C, Spinnler H. 1995. Dual-task paradigm: A means to examine the central executive. *Annals of the New York Academy of Science* **769**: 161-171.

88. Della Sala S, Logie RH. 2001. Theoretical and practical implications of dual-task performance in Alzheimer's Disease. *Brain* **124**: 1479-1481.

89. Dierckx E, Engelborghs S. 2007. Differentiation between mild cognitive impairment, Alzheimer's disease and depression by means of cued recall. *Psychological Medicine* **37**(5): 747-755.

90. Dierckx E, Engelborghs S, De Raedt R, De Deyn PP, Ponjaert-Kristoffersen I. 2007. Mild cognitive impairment: What's in a name? *Gerontology* **53**(1): 28-35.

91. Dubois B, Feldman HH, Jacova C, DeKosky ST, Barberger-Gateau P, Cummings J, Delacourte A, Galasko D, Gauthier S, Jicha G, Meguro K, O'Brien J, Pasquier F, Robert P, Rossor M, Salloway S, Stern Y, Visser PJ, Schetens P. 2007. Reserach criteria for the diagnosis of AD: revising the NINCDS-ADRDA criteria. *The Lancet Neurology* **6**(8): 734-746.

92. Dudas RB, Berrios GE, Hodges JR. 2005a. The Addenbrooke's Cognitive Examination (ACE) in the Differential Diagnosis of Early Dementias Versus Affective Disorder. *Am J Geriatr Psychiatry* **13**(3): 218-226.

93. Dudas RB, Clague F, Thompson SA, Graham KS, Hodges JR. 2005b. Episodic and semantic memory in mild cognitive impairment. *Neuropsychologia* **43**(9): 1266-1276.

94. Duff K, Beglinger LJ, Schultz SK, Moser DJ, McCaffrey RJ, Haase RF, Westervelt HJK, Langbehn DR, Paulsen JS. 2007. Practice effects in the prediction of long-term cognitive outcome in three patient samples: a novel prognostic index. *Archives of Clinical Neuropsychology* **22**: 15-24.

95. Duong A, Whitehead V, Hanratty K, Chertkow H. 2006. The nature of lexico-semantic processing deficits in mild cognitive impairment. *Neuropsychologia* **44**(10): 1928-1935.

96. Economou A, Papageorgiou SG, Karageorgiou C, Vassilopoulos D. 2007. Nonepisodic memory deficits in amnestic MCI. *Cognitive and Behavioral Neurology* **20**(2): 99-106.

97. Egerhazi A, Berecz R, Bartok E, Degrell I. 2007. Automated neuropsychological test battery (CANTAB) in mild cognitive impairment and in alzheimer's disease. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* **31**: 746-751.

98. Estevez-Gonzalez A, Garcia-Sanchez C, Boltes A, Otermin P, Pascual-Sedano B, Gironell A, Kulisevsky J. 2004. Semantic Knowledge of Famous People in Mild Cognitive Impairment and Progression to Alzheimer's Disease. *Dement Geriatr Cogn Disord* **17**(3): 188-195.

99. Fama R, Sullivan EV, Shear PK. 1998. Fluency performance patterns in Alzheimer's disease and Parkinson's disease. *The Clinical Neuropsychologist* **12**: 487-499.

100. Farias ST, Mungas D, Reed BR, Harvey D, Cahn-Weiner D, DeCarli C. 2006. MCI is associated with deficits in everyday functioning. *Alzheimer Disease & Associated Disorders* **20**(4): 217-223.

101. Field A. 2005. *Discovering Statistics Using SPSS*. Sage Publications: London.

102. Fischer JS, Schweizer TA, Atkins JH, Bozanovic R, Norris M, Herrmann N, Nisenbaum R, Rourke SB. 2008. Neurocognitive profiles in older adults with and without major depression. *International Journal of Geriatric Psychiatry* **23**: 851-856.

103. Flanagan JL, Jackson ST. 1997. Test-retest reliability of three aphasia tests: Performance of non-brain-damaged older adults. In 33-43.

104. Folstein M, Folstein M, McHugh P. 1975. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* **12**: 189-198.

105. Folstein MF, Folstein SE, McHugh PR, Fanjiang G. 2001. Mini-mental state examination user's guide. In Odessa F (ed). Psychological Assessment Resources.

106. Fossati P, Harvey PO, LeBastard G, Ergis AM, Jouvent R, Allilaire JF. 2004. Verbal memory performance of patients with a first depressive episode and patients with unipolar and bipolar recurrent depression. *Journal of Psychiatric Research* **38**: 137-144.

107. Fowler KS, Saling MM, Conway EL, Semple JM, Louis WJ. 1995. Computerized delayed matching to sample and paired associate performance in the early detection of dementia. *Applied Neuropsychology* **2**: 72-78.

108. Fowler KS, Saling MM, Conway EL, Semple JM, Louis WJ. 1997. Computerised neuropsychological tests in the early detection of dementia: Prospective findings. *Journal of the International Neuropsychological Society* **3**: 139-146.

109. Fowler KS, Saling MM, Conway EL, Semple JM, Louis WJ. 2002. Paired associate performance in the early detection of DAT. *Journal of the International Neuropsychological Society* **8**: 58-71.

110. Fox NC, Warrington EK, Seiffer AL, Agnew SK, Rossor MN. 1998. Presymptomatic cognitive deficits in individuals at risk of familial Alzheimer's disease. *Brain* **121**: 1631-1639.

111. Fray PJ, Robbins RW, Sahakian BJ. 1996. Neuropsychiatric applications of CANTAB. *International Journal of Geriatric Psychiatry* **11**: 329-336.

112. Galton CJ, Erzinclioglu S, Sahakian BJ, Antoun N, Hodges JR. 2005. A comparison of the Addenbrooke's Cognitive Examination (ACE), conventional neuropsychological assessment and simple MRI-based medial temporal lobe evaluation in the early diagnosis of Alzheimer's Disease. *Cognitive and Behavioral Neurology* **18**(3): 144-150.

113. Geslani DM, Tierney MC, Herrmann N, Szalai JP. 2005. Mild Cognitive Impairment: An Operational Definition and Its Conversion Rate to Alzheimer's Disease. *Dement Geriatr Cogn Disord* **19**(5-6): 383-389.

114. Godbolt AK, Cipolotti L, Anderson VM, Archer H, Janssen JC, Price S, Rossor M, Fox NC. 2005. A decade of pre-diagnostic assessment in a case of familial Alzheimer's disease: tracking progression from asymptomatic to MCI and dementia. *Neurocase* **11**(1): 56-64.

115. Godbolt AK, Cipolotti L, Watt H, Fox NC, Janssen JC, Rossor MN. 2004. The natural history of Alzheimer's disease. A longitudinal presymptomatic study of a family cohort. *Arch Neurol* **61**(11): 1743-1748.

116. Gould RL, Bronw RG, Owen AM, Bullmore ET, Williams SCR, Howard RJ. 2005. Functional neuroanatomy of successful paired associate learning in Alzheimer's disease. *Am J Psychiatry* **162**: 2049-2060.

117. Grabowski TJ. 2008. More to MCI than meets the eye. *Cortex* **44**: 753-756.

118. Graham KS. 2004. GFT normative data. In .

119. Graham NL, Emery T, Hodges JR. 2004. Distinctive cognitive profiles in Alzheimer's disease and subcortical vascular dementia. *J Neurol Neurosurg Psychiatry* **75**(1): 61-71.

120. Greenaway MC, Lacritz LH, Binegar D, Weiner MF, Lipton A, Munro Cullum C. 2006. Patterns of Verbal Memory Performance in Mild Cognitive Impairment, Alzheimer Disease, and Normal Aging. *Cogn Behav Neurol* **19**(2): 79-84.

121. Greene JD, Hodges JR, Baddeley A. 1995. Autobiographical memory and executive function in early dementia of Alzheimer type. *Neuropsychologia* **33**: 1647-1670.

122. Greene JDW, Baddeley AD, Hodges JR. 1996. Analysis of the episodic memory deficit in early Alzheimer's disease: Evidence from the doors and people test. *Neuropsychologia* **34**(6): 537-551.

123. Grober E, Kawas C. 1997. Learning and Retention in Preclinical and Early Alzheimer's Disease*1. *Psychology and Aging* **12**(1): 183-188.

124. Grundman M, Petersen RC, Ferris SH, Thomas RG, Aisen PS, Bennett DA, Foster NL, Jack CR, Jr., Galasko DR, Doody R, Kaye J, Sano M, Mohs R, Gauthier S, Kim HT, Jin S, Schultz AN, Schafer K, Mulnard R, van Dyck CH, Mintzer J, Zamrini EY, Cahn-Weiner D, Thal LJ. 2004. Mild Cognitive Impairment Can Be Distinguished From Alzheimer Disease and Normal Aging for Clinical Trials. *Arch Neurol* **61**(1): 59-66.

125. Gualtieri CT, Johnson LG. 2005. Neurocognitive testing supports a broarder concept of mild cognitive impairment. *American Journal of Alzheimer's Disease and Other Dementias* **20**(359): 366.

126. Guarch J, Marcos T, Salamero M, Blesa R. 2004. Neuropsychological markers of dementia in patients with memory complaints. *International Journal of Geriatric Psychiatry* **19**(4): 352-358.

127. Hachinski V. 2008. Shifts in thinking about dementia. *Journal of the American Medical Association* **300**(18): 2172-2173.

128. Harrison E, Buxton P, Husain M, Wise R. 2000. Short test of semantic and phonological fluency: Normal performance, validity and test-retest reliability. *British Journal of Clinical Psychology* **39**(2): 181-192.

129. Hartley AA, Little DM. 1999. Age related differences and similarities in dual-task interference. *Journal of Experimental Psychology : General* **128**: 416-449.

130. Hasher L, Zacks RT. 1979. Automatic and effortful processes in memory. *Journal of Experimental Psychology : General* **108**: 356-388.

131. Henry JD, Crawford JR. 2005. A meta-analytic review of verbal fluency deficits in depression. *Journal of Clinical and Experimental Neuropsychology* **27**: 78-101.

132. Henry JD, Crawford JR, Phillips LH. 2004. Verbal fluency performance in dementia of the Alzheimer's type: a meta-analysis. *Neuropsychologia* **42**: 1212-1222.

133. Herrmann LL, Goodwin G, Ebmeier K. 2007. The cognitive neuropsychology of depression in the elderly. *Psychological Medicine* **37**(12): 1693-1702.

134. Hodges JR, Erzinclioglua S, Patterson K. 2006. Evolution of Cognitive Deficits and Conversion to Dementia in Patients with Mild Cognitive Impairment: A Very-Long-Term Follow-Up Study. *Dement Geriatr Cogn Disord* **21**: 380-391.

135. Hodges JR, Patterson K. 1995. Is semantic memory consistently impaired early in the course of Alzheimer's disease? Neuroanatomical and diagnostic implications. *Neuropsychologia* **33**(4): 441-459.

136. Hodges JR, Salmon DP, Butters N. 1993. Recognition and naming of famous faces in Alzheimer's disease: A cognitive analysis. *Neuropsychologia* **31**(8): 775-788.

137. Hogervorst E, Combrinck M, Lapuerta P, Rue J, Swales K, Budge M. 2002. The Hopkins Verbal Learning Testing and screening for Dementia. *Dement Geriatr Cogn Disord* **13**(1): 13-20.

138. Holm S. 1979. A simple sequential rejective multiple test proceedure. *Scandinavian Journal of Statistics* **6**: 65-70.

139. Holtzer R, Burright RG, Donovick PJ. 2004. The sensitivity of dual task performance to cognitive status in aging. *Journal of the International Neuropsychological Society* **10**: 230-238.

140. Hudon C, Belleville S, Souchay C, Gely-Nargeot M, Chertkow H. 2006. Memory for gist and detail information in Alzheimer's disease and Mild Cognitive Impairment. *Neuropsychology* **20**(5): 566-577.

141. Jak AJ, Bondi MW, Delano-Wood L, Wierenga C, Corey-Bloom J, Salmon DP, Delis DC. 2009. Quantification of five neuropsychological approaches to defining mild cognitive impairment. *Am J Geriatr Psychiatry* **17**: 368-375.

142. Jefferies E, Lambon Ralph MA. 2006. Semantic impairment in stroke aphasia versus semantic dementia: a case-series comparison. *Brain* : 1-16.

143. Jicha GA. 2008. Clinical features of MCI differ in research and tertiary clinic settings. *Dement Geriatr Cogn Disord* **26**(2): 187-192.

144. Jones S, Laukka EJ, Backman L. 2006. Differential verbal fluency deficits in the preclinical stages of Alzheimer's disease and vascular dementia. *Cortex* **42**(3): 347-355.

145. Joubert S, Felician O, Barbeau EJ, Didic M, Poncet M, Ceccaldi M. 2008. Patterns of semantic memory impairment in mild cognitive impairment. *Behavioural Neurology* **19**(1-2): 35-40.

146. Kalbe E, Salmon E, Perani D, Holthoff V, Sorbi S, Elsner A, Weisenbach S, Brand M, Lenz O, Kessler J, Luedecke S, Ortelli P, Herholz K. 2005. Anosognosia in very mild Alzheimer's disease but not in mild cognitive impairment. *Dement Geriatr Cogn Disord* **19**(5-6): 349-356.

147. Kaplan E, Goodglass H, Weintraub S. 1983. *Boston Naming Test.* Lee & Febiger: Philadelphia.

148. Kendell RE. 1989. Clinical validity. *Psychological Medicine* **19**: 45-55.

149. Knapp M, Prince M. 2007. Dementia UK; summary of key findings. In Alzheimer's Society: London; 1-16.

150. Kohler S, Thomas AJ, Barnett NA, O'Brien JT. 2009. The pattern and course of cognitive impairment in late-life depression. In .

151. Kramer JH, Nelson A, Johnson JK, Yaffe K, Glenn S, Rosen HJ, Miller BL. 2006. Multiple cognitive deficits in amnestic mild cognitive impairment. *Dementia & Geriatric Cognitive Disorders* **22**(4): 306-311.

152. Lafleche G, Albert MS. 1995. Executive Function Deficits in Mild Alzheimer's Disease*1. *Neuropsychology* **9**(3): 313-320.

153. Lam LCW, Ho P, Lui VWC, Tam CWC. 2006. Reduced semantic fluency as an additional screening tool for subjects with questionable dementia. *Dement Geriatr Cogn Disord* **22**(2): 159-164.

154. Lambon R, Patterson K, Graham N, Dawson K, Hodges JR. 2003. Homogeneity and heterogeneity in mild cognitive impairment and Alzheimer's disease: a cross-sectional and longitudinal study of 55 cases. *Brain* **126**(11): 2350-2362.

155. Larner AJ. 2005. An audit of the Addenbrooke's Cognitive Examination (ACE) in clinical practice. *Int J Geriatr Psychiatry* **20**(6): 593-594.

156. Larner AJ. 2006. Audit of the Addenbrooke's Cognitive Examination (ACE) in clinical practice: 2. longitudinal change. *Int J Geriatr Psychiatry* **21**(7): 698-699.

157. Lawton MP, Brody EM. 1969. Assessment of Older People - Self-Maintaining and Instrumental Activities of Daily Living. *Gerontologist* **9**(3P1): 179-&.

158. Lee ACH, Rahman S, Hodges JR, Sahakian BJ, Graham KS. 2003. Associative and recognition memory for novel objects in dementia: implications for diagnosis. *European Journal of Neuroscience* **18**(6): 1660-1670.

159. Lee DY, Youn JC, Choo IH, Kim KW, Jhoo JH, Pak Y, Suh KW, Woo JI. 2006. Combination of Clinical and Neuropsychologic Information as a Better Predictor of the Progression to Alzheimer Disease in Questionable Dementia Individuals. *Am J Geriatr Psychiatry* **14**(2): 130-138.

160. Lehrner J, Gufler R, Guttmann G, Maly J, Gleib A, Auff E, Dal-Bianco P. 2005. Annual conversion to Alzheimer disease among patients with memory

complaints attending an outpatient memory clinic: the influence of amnestic mild cognitive impairment and the predictive value of neuropsychological testing. *The Middle European Journal of Medicine* (117): 18-629.

161. Lemelin S, Baruch P. 1998. Clinical psychomotor retardation and attention in depression. *Journal of Psychiatric Research* **32**: 81-88.

162. Levinoff EJ, Verret L, Akerib V, Phillips NA, Babins L, Kelner N, Chertkow H. 2006. Cognitive estimation impairment in Alzheimer's disease and mild cognitive impairment. *Neuropsychology* **20**(1): 123-132.

163. Lezak MD. 2004. *Neuropsychological Assessment*. Oxford University Press: London.

164. Lichtenberg PA, Ross R, Millis SR, Manning CA. 1995. The relationship between depression and cognition in older adults: A cross-validation study. *J Gerontol B Psychol Sci Soc Sci* **50B**(25): 32.

165. Loewenstein DA, Acevedo A, Agron J, Duara R. 2007a. Stability of neurocognitive impairment in different subtypes of mild cognitive impairment. *Dement Geriatr Cogn Disord* **23**(2): 82-86.

166. Loewenstein DA, Acevedo A, Agron J, Issacson R, Strauman S, Crocco E, Barker WW, Duara R. 2006a. Cognitive profiles in alzheimer's disease and in mild cognitive impairment of different etiologies. In 309-315.

167. Loewenstein DA, Acevedo A, Argon J, Martinez G, Duara R. 2007b. The use of amnestic and nonamnestic composite measures at different thresholds in the neuropsychological diagnosis of MCI. *Journal of Clinical and Experimental Neuropsychology* **29**(3): 300-307.

168. Loewenstein DA, Acevedo A, Ownby R, Agrot J, Barker WW, Isaacson R, Strauman S, Duara R. 2006b. Using different memory cutoffs to assess mild cognitive impairment. *Am J Geriatr Psychiatry* **14**(11): 911-919.

169. Logie RH, Cocchini G, Della Sala S, Baddeley AD. 2004. Is there a specific executive capacity for dual task coordination? Evidence from Alzheimer's disease. *Neuropsychology* **18**(3): 504-513.

170. Lonie JA, Hermann L, Tierney KM, Donaghey C, O'Carroll R, Lee A, Ebmeier KP. 2009a. Lexical and semantic fluency discrepancy scores in aMCI and early Alzheimer's disease. *Journal of Neuropsychology* **3**: 79-92.

171. Lonie JA, Herrmann LL, Donaghey CL, Ebmeier KP. 2008. Clinical referral patterns and cognitive profile in mild cognitive impairment. *British Journal of Psychiatry* **192**(1): 59-64.

172. Lonie JA, Tierney KM, Herrmann LL, Donaghey C, O'Carroll RE, Lee A, Ebmeier KP. 2009b. Dual task performance in early Alzheimer's disease, amnestic mild cognitive impairment and depression. *Psychological Medicine* **39**: 23-31.

173. Lopez MN, Charter RA, Mostafavi B, Nibut LP, Smith WE. 2005. Psychometric properties of the Folstein Mini-Mental State Examination. *Assessment* **12**(2): 137-144.

174. Lopez-Pousa S, Turon-Estrada A, Garre-Olmo J, Pericot-Nierga I, Lozano-Gallego M, Vilalta-Franch M, Hernandez-Ferrandiz M, Morante-Munoz V, Isern-Vila A, Gelada-Batlle E, Majo-Llopart J. 2005. Differential efficacy of treatment with acetylcholinesterase inhibitors in patients with mild and moderate Alzheimer's disease over a 6-month period. *Dement Geriatr Cogn Disord* **19**: 189-195.

175. Lowndes G, Savage G. 2007. Early detection of memory impairment in alzheimer's disease: a neurocognitive perspective on assessment. *Neuropsychology Review* **17**: 193-202.

176. Lu PH, Edland SD, Teng E, Ingus K, Petersen RC, Cummings JL. 2009. Donepezil delays progression to AD in MCI subjects with depressive symptoms. *Neurology* **72**: 2115-2121.

177. Lu Z, Neuse J, Madigan S, Dosher BA. 2005. Fast decay of iconic memory in observers with mild cognitive impairments. *PNAS* **102**(1): 1797-1802.

178. Lucas JA, Ivnik RJ, Smith GE. 1998. Mayo's Older Americans Normative Studies: Category fluency norms. *Journal of Clinical and Experimental Neuropsychology* **20**: 194-200.

179. MacPherson SE, Della Sala S, Logie RH. 2004. Dual-task interference of encoding and retrieval processes in healthy and impaired working memory. *Cortex* **40**(1): 183-184.

180. Mahurin RK, Velligan DI, Hazleton B, Davis JM, Eckert S, Miller AL. 2006. Trail making test errors and executive function in schizophrenia and depression. *Clinical Neuropsychology* **20**: 271-288.

181. Marcos A, Gil P, Barabash A, Rodrigues R, Encinas M, Fernandez C, Cabranes JA. 2006. Neuropsychological markers of progression from mild cognitive impairment to Alzheimer's disease. *American Journal of Alzheimer's Disease and Other Dementias* **21**(3): 189-196.

182. Marczinski CA, Kertesz A. 2006. Category and letter fluency in semantic dementia, primary progressive aphasia, and Alzheimer's disease. *Brain and Language* **97**(3): 258-265.

183. Mathuranath PS, Nestor PJ, Berrios GE, Rakowicz W, Hodges JR. 2000. A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. *Neurology* **55**(11): 1613-1620.

184. Mattis S. 1988. Dementia Rating Scale (DRS). In Psychological Assessment Resources, Inc. : Odessa: FL.

185. McCaffrey RJ, Ortega A, Orsillo SM. 1992. Practice effects in repeated neuropsychological assessments. *The Clinical Neuropsychologist* **6**: 32-42.

186. McKeith IG, Perry EK, Perry RH. 1999. Report of the second dementia with lewy body international workshop : diagnosis and treatment. *Neurology* **53**: 902.

187. McKenna P, Warrington EK. 1980. Testing for nominal dysphasia. *J Neurol Neurosurg Psychiatry* **43**(9): 781-788.

188. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. 1984. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**(7): 939-944.

189. Mebane-Sims I. 2009. 2009 Alzheimer's disease facts and figures. *Alzheimer's and Dementia* **5**(3): 234-270.

190. Mendez MF, Ghajarania M, Perryman KM. 2009. Posterior cortical atrophy: clinical characteristics and differences compared to Alzheimer's disease. *Dement Geriatr Cogn Disord* **14**: 33-40.

191. Mesulam MM. 2001. Primary progressive aphasia. *Ann Neurol* **49**: 425-432.

192. Meyer JS, Xu G, Thornby J, Chowdhury M, Quach M. 2002. Longitudinal analysis of abnormal domains comprising mild cognitive mipairment (MCI) during aging. *Journal of the Neurological Sciences* **201**: 19-25.

193. Meyers JE, Meyers KR. 1995. Rey Complex Figure Test under four different administration procedures. *The Clinical Neuropsychologist* **9**: 63-67.

194. Mickes L, Wixted JT, Fennema-Notestine C, Galasko D, Bondi M, Thal LJ, Salmon DP. 2007. Progressive impairment on neuropsychological tasks in a longitudinal study of preclinical alzheimer's disease. *Neuropsychology* **21**(6): 696-705.

195. Mioshi E, Dawson K, Mitchell J, Arnold R, Hodges JR. 2006. The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *International Journal of Geriatric Psychiatry* **21**(11): 1078-1085.

196. Mirra SS, Heyman A, McKeel D, Sumi SM, Crain BJ, Brownlee LM. 2009. The consortium to establish a registry for Alzheimer's disease (CERAD). Part II: Starndardization of the neuropathologic assessment of Alzheimer's disease. *Neurology* **41**: 479-486.

197. Mitchell T, Woodward M, Hirose Y. 2008. A survey of attitudes of clinicians towards the diagnosis and treatment of mild cognitive impairment in Australia and New Zealand. *International Psychogeriatrics 20(1)()(pp 77-85), 2008 Date of Publication: May 2008* (1): 77-85.

198. Mitrushina M, Satz P. 1991. Effect of repeated administration of a neuropsychological battery in the elderly. *Journal of Clinical Psychology* **47**: 790-801.

199. Mitrushina M, Satz P. 1995. Repeated testing of normal elderly with the boston naming test. *Aging* **7**: 123-127.

200. Mitrushina MN, Boone KB, D'Elia LF. 1999. *Handbook of normative data for neuropsychological assessment*. Oxford University Press: New York.

201. Morris JC, Heyman A, Mohs RC. 1989. The consortium to establish a registry for Alzheimer's disease (CERAD), part I: clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* **39**: 1159-1165.

202. Morris R, Evenden J, Sahakian B, Robbins T. 1987. Computer aided assessment of dementia: Comparative studies of neuropsychological deficits in Alzheimer type dementia and Parkinson's disease. In *Cognitive neurochemistry*, Stahl.S., Iversen S, Goodman E (eds). Oxford University Press: Oxford, UK; 21-36.

203. Murphy KI, Rich JB, Troyer AK. 2006. Verbal fluency patterns in amnestic mild cognitive impairment are characteristic of Alzheimer's type dementia. *Journal of the International Neuropsychological Society* **12**(4): 570-574.

204. Nathan J, Wilkinson D, Stammers S, Low JL. 2001. The role of tests of frontal executive functioning in the detection of mild dementia. *International Journal of Geriatric Psychiatry* **16**: 18-26.

205. National Collaborating Centre for Mental Health. 2006. Nice clinical guideline 42. In .

206. Neary D, Snowden JS, Gustafson L, Passant U, Stuss D, Black S, Freedman M, Kertesz A, Robert PH, Albert M, Boone K, Miller BL, Cummings J, Benson DF. 1998. Frontotemporal lobar degeneration. A consensus on clinical diagnostic criteria. *Neurology* **51**(1546): 1554.

207. Nebes RD, Butters MA, Houck PR, Zmuda MD, Aizenstein H, Pollock BG, Mulsant BH, Reynolds CF. 2001. Dual-task performance in depressed geriatric patients. *Psychiatry Research* **102**: 139-151.

208. Nelson H, Willison J. 1991. *National Adult Reading Test (NART).* Windsor, London.

209. Nordahl CW, Ranganath C, Yonelinas AP, DeCarli C, Reed BR, Jagust WJ. 2005. Different mechanisms of episodic memory failure in mild cognitive impairment. *Neuropsychologia* **43**(11): 1688-1697.

210. Nordlund A, Rolstad S, Hellstrom P, Sjogren M, Hansen S, Wallin A. 2005. The Goteborg MCI study: mild cognitive impairment is a heterogeneous condition. *J Neurol Neurosurg Psychiatry* **76**(11): 1485-1490.

211. Nordlund A, Rolstad S, Klang O, Lind K, Hansen S, Wallin A. 2007. Cognitive profiles of mild cognitive impairment with and without vascular disease. *Neuropsychology* **21**(6): 706-712.

212. Nutter-Upham KE, Saykin AJ, Rabin LA, Roth RM, Wishart HA, Pare N, Flashman LA. 2008. Verbal fluency performance in amnestic MCI and older adults with cognitive complaints. *Archives of Clinical Neuropsychology* **23**: 229-241.

213. O'Bryant SE, Humphreys JD, Smith GE, Ivnik RJ, Graff-Radford RN, Petersen RC, Lucas JA. 2008. Detecting dementia with the MMSE in highly educated individuals. *Arch Neurol* **65**(7): 963-967.

214. O'Carroll RE, Curran SM, Ross M. 1994. The differentiation of major depression from dementia of the Alzheimer type using within-subject neuropsychological discrepancy analysis. *British Journal of Clinical Psychology* **33**: 23-32.

215. O'Connell H, Coen R, Kidd N, Warsi M, Chin A, Lawlor B. 2004. Early detection of Alzheimer's disease (AD) using the CANTAB Paired Associates Learning Test. *International Journal of Geriatric Psychiatry* **19**: 1207-1208.

216. Palmer K, Backman L, Winblad B, Fratiglioni L. 2003. Detection of Alzheimer's disease and dementia in the preclinical phase: population based cohort study. *BMJ* **326**(7383): 245.

217. Palmer K, Berger AK, Monastero R, Winblad B, Backman L, Fratiglioni L. 2007. Predictors of progression from mild cognitive impairment to Alzheimer disease. *Neurology 68(19)()(pp 1596-1602), 2007 Date of Publication: May 2007* (19): 1596-1602.

218. Panza F, D'Introno A, Colacicco AM, Capurso C, Del Parigi A, Caselli RJ, Pilotto A, Argentieri G, Scapicchio PL, Scafato E, Capurso A, Solfrizzi V. 2005. Current Epidemiology of Mild Cognitive Impairment and Other Predementia Syndromes. *Am J Geriatr Psychiatry* **13**(8): 633-644.

219. Parasuraman R, Haxby JV. 1993. Attention and brain function in Alzheimer's disease: A review. *Neuropsychology* **7**(3): 242-272.

220. Perri R, Serra L, Carlesimo GA, Caltagirone C. 2007. Amnestic mild cognitive impairment: Difference of memory profile in subjects who converted or did not convert to Alzheimer's disease. *Neuropsychology* **21**(5): 549-558.

221. Perrotin A, Belleville S, Isingrini M. 2007. Metamemory monitoring in mild cognitive impairment: Evidence of a less accurate episodic feeling-of-knowing. *Neuropsychologia 45(12)()(pp 2811-2826), 2007 Date of Publication: 2007* (12): 2811-2826.

222. Perry RJ, Hodges JR. 2000. Fate of Patients with Questionable (Very Mild) Alzheimer's Disease: Longitudinal Profiles of Individial Subjects' Decline. *Dement Geriatr Cogn Disord* **11**: 342-349.

223. Perry RJ, Watson P, Hodges JR. 2000. The nature and staging of attention dysfunction in early (minimal and mild) Alzheimer's disease: relationship to episodic and semantic memory impairment. *Neuropsychologia* **38**(3): 252-271.

224. Perry RJ, Hodges JR. 1999. Attention and executive deficits in Alzheimer's disease: A critical review. *Brain* **122**(3): 383-404.

225. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Kokmen E. 1999. Mild Cognitive Impairment: Clinical Characterization and Outcome. *Archive of Neurology* **56**: 303-308.

226. Petersen RC. 2004. Mild Cognitive Impairment as a diagnostic entity. *Journal of International Medicine* **256**: 183-194.

227. Petersen RC. 2005a. Mild Cognitive Impairment as a Clinical Entity and Treatment Target. *Archive of Neurology* **62**(7): 1160-1163.

228. Petersen RC. 2005b. Mild cognitive impairment: Useful or not? *Alzheimer's and Dementia* **1**: 5-10.

229. Petersen RC. 2005c. Mild Cognitive impairment: Where are we? *Alzheimer Dis Assoc Disord* **19**: 166-169.

230. Petersen RC, O'Brien J. 2006. Mild cognitive impairment should be considered for DSM-V. *J Geriatr Psychiatry Neurol* **19**(3): 147-154.

231. Petersen RC, Smith GE, Ivnik RJ, Kokmen E, Tangalos EG. 1994. Memory function in very early Alzheimer's disease. *Neurology* **44**(5): 867-872.

232. Petersen RC, Thomas RG, Grundman M, Bennett D, Doody R, Ferris S, Galasko D, Jin S, Kaye J, Levey A, Pfeiffer E, Sano M, Van Dyck CH, Thal LJ. 2005. Vitamin E and donepezil for the treatment of mild cognitive impairment. *N Engl J Med* **352**(23): 2379-2388.

233. Petersen RC, Doody R, Kurz A, Mohs RC, Morris JC, Rabins PV, Ritchie K, Rossor M, Thal L, Winblad B. 2001. Current Concepts in Mild Cognitive Impairment. *Arch Neurol* **58**(12): 1985-1992.

234. Pike EK, Savage G. 2008. Memory profiling in mild cognitive impairment: can we determine risk for Alzheimer's disease? *Journal of Neuropsychology* **2**(2): 361-372.

235. Pike KE, Rowe CC, Moss SA, Savage G. 2008. Memory profiling with paired associate learning in Alzheimer's disease, mild cognitive impairment and healthy aging. *Neuropsychology* **22**(6): 718-728.

236. Pioggiosi PP, Berardi D, Ferrari B, Quartesan R, De Ronchi D. 2006. Occurrence of cognitive impairment after age 90: MCI and other broadly used concepts. *Brain Research Bulletin* **68**(4): 227-232.

237. Popoola A, O'Donovan M, Mullan E. 2008. An assessment of the new memory clinic at Cork University Hospital. *Irish Journal of Psychological Medicine* **25**(2): 61-65.

238. Portet F, Ousset PJ, Visser PJ, Frisoni GB, Nobili J, Scheltens P, Vellas B, Touchon J. 2006. Mild Cognitive Impairment in medical practice: critical review of the concept and new diagnostic procedure. Report of the MCI working group of the European Consortium on Alzheimer's disease (EADC). *J Neurol Neurosurg Psychiatry* **77**(6): 714-718.

239. Rabbitt PMA. 1997. *Methodology of frontal and executive function*. Psychology Press: Hove.

240. Rajji TK, Miranda D, Mulsant BH, Lotz M, Houck P, Zmuda MD, Bensasi S, Reynolds CF, Butters MA. 2009. The MMSE is not an adequate screening cognitive instrument in studies of late-life depression. *Journal of Psychiatric Research* **43**: 464-470.

241. Rapp MA, Dahalman K, Sano M, Grossman HT, Haroutunian V, Gorman JM. 2005. Neuropsychological differences between late-onset and recurrent geriatric major depression. *Am J Psychiatry* **162**: 691-698.

242. Rasmusson DX, Bylsma FW, Brandt J. 1995. Stability of performance on the Hopkins Verbal Learning Test. *Archives of Clinical Neuropsychology* **10**(1): 21-26.

243. Ravaglia G, Forti P, Maioli F, Servadei L, Martelli M, Brunetti N, Bastagli L, Mariani E. 2005. Screening for mild cognitive impairment in elderly ambulatory patients with cognitive complaints. *Aging - Clinical and Experimental Research* (5): 374-379.

244. Ravdin LD, Katzen HL, Agrawal P, Relkin NR. 2003. Letter and Semantic Fluency in Older Adults: Effects of Mild Depressive Symptoms and Age-Stratified Normative Data. *The Clinical Neuropsychologist* **17**(2): 195-202.

245. Reitan RM. 1985. *Halstead-Reitan Neuropsychological Test Battery*. Reitan Neuropsychology Laboratory/Press: Tuscon.

246. Rentz DM, Huh TJ, Faust RR, Budson AE, Scinto LFM, Sperling RA, Daffner KR. 2004. Use of IQ-adjusted norms to predict progressive cognitive decline in highly intelligent older individuals. *Neuropsychology* **18**(1): 38-49.

247. Rey A. 1941. Psychological examination of a case of post-traumatic encephalopathy. *Archieves de Psychologie* **28**: 286-340.

248. Riberio F, deMendonca A, Guerreiro M. 2006. Mild Cognitive Impairment: Deficits in cognitive domains other than memory. *Dement Geriatr Cogn Disord* **21**: 284-290.

249. Ritchie K, Artero S, Touchon J. 2001. Classification criteria for mild cognitive impairment: A population-based validation study. *Neurology* **56**(1): 37-42.

250. Ritter E, Despres O, Monsch AU, Manning L. 2006. Topographical recognition memory sensitive to amnestic mild cognitive impairment but not to depression. *International Journal of Geriatric Psychiatry* **21**(10): 924-929.

251. Robbins TW, James MO, Sahakian BJ, Lawrence AD, McInnes L. 1994. A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: Implications for theories of executive functioning and cognitive aging. *Journal of the International Neuropsychological Society* **4**: 474-490.

252. Rose SE, McMahon KL, Janke AL, O'Dowd B, de Zubicaray G, Strudwick MW, Chalk JB. 2006. Diffusion indices on magnetic resonance imaging and neuropsychological performance in amnestic mild cognitive impairment. *J Neurol Neurosurg Psychiatry* **77**(10): 1122-1128.

253. Sager MA, Hermann BP, La Rue A, Woodard JL. 2006. Screening for dementia in community-based memory clinics. *Wisconsin Medical Journal* **105**(7): 25-29.

254. Salmon DP, Bondi MW. 2009. Neuropsychological assessment of dementia. *Annual Review of Psychology* **60**: 257-282.

255. Salthouse TA. 1996. The processing-speed theory of adult age differences in cognition. *Psychol Rev* **103**(3): 403-428.

256. Saxton J, Ratcliff G, Munro CA, Coffey EC, Becker JT, Fried L, Kuller L. 2000. Normative Data on the Boston Naming Test and Two Equivalent 30-Item Short Forms. *The Clinical Neuropsychologist* **14**(4): 526-534.

257. Scahill VL, Hodges JR, Graham KS. 2005. Can episodic memory tasks differentiate semantic dementia from Alzheimer's Disease? *Neurocase* **11**(6): 441-451.

258. Schmand B, Huizenga HM, van Gool WA. 2010. Meta-analysis of CSF and MRI biomarkers for detecting prelcinical Alzheimer's disease. *Psychological Medicine* **40**: 135-145.

259. Schmidtke K, Hermeneit S. 2007. High rate of conversion to Alzheimer's disease in a cohort of amnestic MCI patients. *International Psychogeriatrics* **20**(1): 96-108.

260. Schrijnemaekers AMC, De Jager CA, Hogervorst E, Budge MM. 2006. Cases with Mild Cognitive Impairment and Alzheimer's Disease Fail to Benefit from Repeated Exposure to Episodic Memory Tests as Compared with Controls. *Journal of Clinical and Experimental Neuropsychology* **28**(3): 438-455.

261. Sebastian MV, Menor J, Elousa MR. 2006. Attentional dysfunction of the central executive in AD: evidence from dual task and perseveration errors. *Cortex* **42**: 1015-1020.

262. Sherman AM, Massman PJ. 1999. Prevalence and Correlates of Category Versus Letter Fluency Discrepancies in Alzheimer's Disease. *Archives of Clinical Neuropsychology* **14**(5): 411-418.

263. Shulman KI, Herrmann N, Brodaty H. 2006. Survey of brief cognitive sceening instruments. *International Psychogeriatrics* **18**: 281-294.

264. Slavin MJ, Sandstrom CK, Tran TT, Doraiswamy PM, Petrella JR. 2007. Hippocampal volume and the mini-mental state examination in the diagnosis of amnestic mild cognitive impairment. *American Journal of Roentgenology* **188**(5): 1404-1410.

265. Smith A. 1982. *Symbol Digits Modalities Test*. Western Psychological Services: Los Angeles.

266. Snow WG, Tierney MC, Zorzitto ML. 1988. One-year test-retest reliability of selected neuropsychological tests in older adults. *Journal of Clinical & Experimental Neuropsychology* **10**: 60.

267. Spreen O, Strauss E. 1998. *A compendium of neuropsychological tests. Administration, norms and commentary.* Oxford University Press: New York.

268. Squire LR. 1992. Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review* **99**(2): 195-231.

269. Starr JM, Lonie JA. 2007. The influence of pre-morbid IQ on Mini-Mental State Examination score at time of dementia presentation. *International Journal of Geriatric Psychiatry* **22**: 382-384.

270. Steffens DC, Potter GG. 2008. Geriatric depression and cognitive impairment. *Psychological Medicine* **38**: 163-175.

271. Stockholm J, Vogel A, Gade A, Waldemar G. 2006. Heterogeneity in Executive Impairment in Patients with Very Mild Alzheimer's Disease. *Dement Geriatr Cogn Disord* **22**(1): 54-59.

272. Storandt M, Grant EA, Miller JP, Morris JC. 2006. Longitudinal course and neuropathologic outcomes in original vs revised MCI and in pre-MCI. *Neurology* **67**(3): 467-473.

273. Swainson R, Hodges JR, Galton CJ, Semple J, Michael A, Dunn BD, Iddon JL, Robbins TW, Sahakian BJ. 2001. Early detection and differential diagnosis of Alzheimer's disease and depression with neuropsychological tasks. *Dement Geriatr Cogn Disord* **12**(4): 265-280.

274. Tabert MH, Albert SM, Borukhova-Milov L, Camacho Y, Pelton G, Liu X, Stern Y, Devanand DP. 2002. Functional deficits in patients with mild cognitive impairment: Prediction of AD. *Neurology* **58**(5): 758-764.

275. Tabert MH, Manly JJ, Liu XH, Pelton GH, Rosenblum S, Jacobs M, Zamora D, Goodkind M, Bell K, Stern Y, Devanand DP. 2006. Neuropsychological prediction of conversion to Alzheimer disease in patients with mild cognitive impairment. *Arch Gen Psychiatry* **63**(8): 916-924.

276. Tang-Wai DF, Knopman DS, Geda YE, Edland SD, Smith GE, Ivnik RJ, Tangalso EG, Boeve BF, Petersen RC. 2003. Comparison of the short test of mental status and the mini-mental state examination in mild cognitive impairment. *Arch Neurol* **60**(1777): 1781.

277. Thomas AJ, Gallagher P, Robinson LF, Porter RJ, Young AH, Ferrier IN, O'Brien JT. 2009. A comparison of neurocognitive impairment in younger and older adults with major depression. *Psychological Medicine* **39**: 725-733.

278. Thomas P, Goudemand M, Rousseaux M. 1999. Attentional resource in major depression. *European Archives of Psychiatry and Clinical Neuroscience* **249**: 79-85.

279. Thompson SA, Graham KS, Patterson K, Sahakian BJ, Hodges JR. 2002. Is Knowledge of Famous People Disproportionately Impaired in Patients With Early and Questionable Alzheimer's Disease?*1. *Neuropsychology* **16**(3): 344-358.

280. Tombaugh TN. 2005. Test-retest reliability coefficients and 5-year change scores for the MMSE and 3MS. *Archives of Clinical Neuropsychology* **20**(4): 485-503.

281. Tombaugh TN. 2004. Trail Making Test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology* **19**(2): 203-214.

282. Tombaugh TN, Kozak J, Rees L. 1999. Normative Data Stratified by Age and Education for Two Measures of Verbal Fluency: FAS and Animal Naming. *Archives of Clinical Neuropsychology* **14**(2): 167-177.

283. Troster AI. 2009. Neuropsychological characteristics of dementia with lewy bodies and parkinson's disease with dementia: differentiation, early detection, and implications for 'mild cognitive impairment' and biomarkers. *Neuropsychology Review* **18**: 103-119.

284. Vanderploeg RD, Schinka JA, Jones T, Small BJ, Borenstein Graves A, Mortimer JA. 2000. Elderly norms for the Hopkins Verbal Learning Test-Revised. *The Clinical Neuropsychologist* **14**(3): 318-324.

285. Visser PJ, Kester A, Jolles J, Verhey F. 2006. Ten-year risk of dementia in subjects with mild cognitive impairment. *Neurology* **67**(7): 1201-1207.

286. Visser PJ, Verhey FJ. 2008. Mild cognitive impairment as predictor for Alzheimer's disease in clinical practice: Effect of age and diagnostic criteria. *Psychological Medicine* (1): 113-122.

287. Vogel A, Gade A, Stokholm J, Waldemar G. 2005. Semantic memory impairment in the earliest phases of Alzheimer's disease. *Dement Geriatr Cogn Disord* **19**: 75-81.

288. Wallenstein GV, Eichenbaum H, Hasselmo ME. 1998. The hippocampus as an associator of discontinguous events. *Trends in Neurosciences* **21**(8): 317-323.

289. Warrington EK, James M. 1991. *The Visual Object and Space Perception Battery*. Thames Valley Test Company: Bury St, Edmunds, UK.

290. Warrington EK. 1997. The graded naming test: A restandardisation. *Neuropsychological Rehabilitation* **7**(2): 143-146.

291. Wechsler D. 1987. *The Wechsler Memory Scale-Revised*. Psychological Corporation: San Antonio.

292. Wechsler D. 1997. *WAIS-III administration and scoring manual*. The Psychological Corporation: San Antonio.

293. Weimer DL, Sager MA. 2009. Early identification and treatment of Alzheimer's disease: Social and fiscal outcomes. *Alzheimers & Dementia* **5**(3): 215-226.

294. Westerberg CE, Paller KA, Mayes AR, Holdstock JS, Reber PJ. 2006. When memory does not fail: familiarity based recognition in mild cognitive impairment and Alzheimer's disease. *Neuropsychology* **20**(2): 193-205.

295. Whalley LJ, Starr JM, Athawes R, Hunter D, Pattie A, Deary IJ. 2000. Childhood mental ability and dementia. *Neurology* **55**(10): 1455-1459.

296. Wimo A, Windbald B, Aguero-Torres H, von Strauss E. 2003. The magnitude of dementia occurrence in the world. *Alzheimer's Disease and Associated Disorders* **17**: 63-67.

297. Winblad B, Palmer K, Kivipelto M, Jelic V, Fratiglioni L, Wahlund LO, Nordberg A, Backman L, Albert M, Almkvist O, Arai H, Basun H, Blennow K, de Leon M, DeCarli C, Erkinjuntti T, Giacobini E, Graff C, Hardy J, Jack C, Jorm A, Ritchie K, van Duijn C, Visser P, Petersen RC. 2004. Mild cognitive impairment - beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *Journal of Internal Medicine* **256**(3): 240-246.

298. Wylie SA, Ridderinkhof KR, Eckerle MK, Manning CA. 2007. Inefficient response inhibition in individuals with mild cognitive impairment. *Neuropsychologia* **45**: 1408-1419.

299. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, Leirer VO. 1983. Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research* **17**: 37-49.

300. Zakzanis KK, Leach L, Kaplan E. 1998. On the nature and pattern of neurocognitive function in major depressive disorder. *Neuropsychiatry, Neuropsychology and Behavioural Neurology* **11**: 111-119.

301. Zec RF, Markweel SJ, Burkett NR, Larsen DL. 2005. A longitudinal study of confrontation naming in the 'normal' elderly. *Journal of the International Neuropsychological Society* **11**: 716-726.

Appendices

Appendix A Papers Arising

Appendix B Tables

Appendix C Figures

Appendix D Measurement Scales

Appendix A Papers Arising

INTERNATIONAL JOURNAL OF GERIATRIC PSYCHIATRY
Int J Geriatr Psychiatry (2009)
Published online in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/gps.2208

REVIEW ARTICLE

Screening for mild cognitive impairment: a systematic review

Jane A. Lonie[1], Kevin M. Tierney[1] and Klaus P. Ebmeier[1,2]*

[1]*Division of Psychiatry University of Edinburgh, Royal Edinburgh Hospital, Edinburgh UK*
[2]*Department of Psychiatry University of Oxford, Warneford Hospital, Oxford, UK*

SUMMARY

Objective Patients with mild cognitive impairment account for a significant number of referrals to old age psychiatry services and specialist memory clinics. The cognitive evaluation of such patients is commonly restricted to brief dementia screens, with no consideration to their suitability for assessing MCI. Here, we review the utility of such cognitive screens for MCI and provide an overview of validated instruments.

Methods We identified papers published after Petersen and colleagues 1999 MCI criteria (Petersen et al., 1999) and examining face-to-face cognitive screening for MCI from publication databases using combinations of the search terms 'mild cognitive impairment' and 'cognitive screening'. We also combined the former search with the names of 39 screening tests recently identified in a relevant review (Cullen et al., 2007).

Results Fifteen cognitive screening instruments were identified, 11 cover a restricted range of cognitive domains. High sensitivity and specificity for MCI relative to healthy controls were reported for two comprehensive and two noncomprehensive screening instruments, adequate test-retest and inter-rater reliability for only one of these. With the exception of three studies, sample sizes were universally small (i.e. $n \leq 100$), and prognostic values were reported for only two of the identified 15 screening measures. Sensitivities of the full domain measures were universally high, but information about their specificity against psychiatric and non-progressive neurological conditions and predictive validity is lacking.

Conclusion Several cognitive screening instruments afford the clinician the ability to detect MCI, early AD, and in some cases non-AD dementia, but they cannot currently be used to make reliable inferences about the course and eventual outcome of MCI.

KEY WORDS — systematic review; screening; mild cognitive impairment; dementia; Alzheimer's disease; vascular dementia; differential diagnosis; neuropsychology

INTRODUCTION

'Mild cognitive impairment' (MCI) is currently the most widely used term to describe cognitive impairment with sound functional capacity and (usually) isolated memory impairment, i.e. without dementia (Petersen et al., 1999, 2001; Mitchell et al., 2008). Although 'multi-domain' and 'non-amnestic' MCI subtypes (Petersen, 2004) have been identified, most MCI studies have focused on the amnestic form (aMCI) (Petersen *et al*. 1999). Petersen's criteria (Petersen et al., 1999, 2001; Petersen, 2004) do not specify the means by which intact other areas of cognitive functioning should be established. The relative rarity of the pure amnestic MCI subtype makes it likely that MCI study samples include a combination of both amnestic single and amnestic multi-domain subtypes (Alladi et al., 2006; Lonie et al., 2008b). In this review, MCI is therefore generally used to describe such single and multi-domain amnestic forms.

MCI is associated with elevated rates of conversion to Alzheimer's disease (Petersen et al., 1999; Busse et al., 2003a; Lehrner et al., 2005; Alladi et al., 2006; Busse et al., 2006; Storandt et al., 2006; Tabert et al., 2006; Palmer et al., 2007). Both neuroimaging (Hirao

*Correspondence to: Prof K. P. Ebmeier, Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford, OX3 7JX, UK. E-mail: klaus.ebmeier@psych.ox.ac.uk

Received 4 August 2008
Accepted 24 November 2008

et al., 2005; Jessen et al., 2006; Rose et al., 2006; Saykin et al., 2006) and neuropathological studies (Morris et al., 2001; Bennett et al., 2005; Albert and Blacker 2006) report intermediate findings between those seen in normal aging and early AD. With the development of disease modifying or halting agents, recognition of AD at the earliest possible stage will become increasingly important.

Estimates for cognitive complaints in the older adult general population range from 3.1% (Busse et al., 2003b) to 15% (Frisoni et al., 2000), depending on case definition and mean age of the sample. The cognitive evaluation of patients with MCI therefore forms an important component of geriatric care (Petersen and O'Brien, 2006) and referrals to specialist memory clinics (Alladi et al., 2006; Lonie et al., 2008b).

In a recent review of cognitive screening in dementia, Cullen and colleagues (2007) caution against a 'one size fits all' approach to dementia screening and advocate the use of a wider range of specialised cognitive screening tools for different situations. In recognition of the different purposes for which cognitive screening measures are used, we divide our review between screening measures that provide an indication of functioning across a comprehensive range of cognitive functions and those assessing a restricted range of cognitive domains. Comprehensive cognitive screening measures are likely to be more suitable for secondary or tertiary specialist settings, while the shorter cognitive screens that do not rely on specialist theoretical knowledge or training are likely to have greater application in screening for cognitive impairment in the community or in a primary care setting.

METHODS

Papers following upon the publication of Petersen and Windblad's MCI criteria (Petersen et al., 1999; Winblad et al., 2004) were identified by electronic databases searches of BIOSIS Previews, Embase, Medline, PsychINFO, Social Sciences Citation Index, ASSIA Plus, IBSS Online, PsychARTICLES, ISI Proceedings and Web of Knowledge (Figure 1). We used combinations of the search terms 'Mild Cognitive Impairment' and 'cognitive screening'. The same databases were searched again combining the search term 'Mild Cognitive Impairment' with the names of 39 cognitive screening tests identified in a recent review of tests for cognitive impairment (Cullen et al., 2007) (Figure 2). Non-screening tests of performance in a single cognitive domain, or their combinations, were excluded as were non face-to-face screening measures and those that are wholly carer or informant-rated (Lines et al., 2003; Giaquinto and Parnetti, 2006; van Uffelen et al., 2007), although the latter may well provide a significant adjunctive or standalone contribution to MCI screening in primary care (Galvin et al., 2007). Where more than one version of a screening instrument existed, we included only the most up-to-date version of the test on the assumption that modification to a previous test version had lead to an improvement in one or more aspects of the test.

Fifty-seven papers were identified. Seven of these were excluded as they were not face-to-face or non-cognitive in nature. The findings were reported in a language other than English for a further three papers. None of the screening measures reported in the non-English papers were reviewed as part of included papers that were written in English. The remaining studies were excluded as they comprised duplicate data ($n = 3$) or did not meet the study objectives i.e. 'to investigate the screening utility of cognitive screening measures in MCI' ($n = 12$). Petersen's MCI criteria (Petersen et al., 1999) were not met, or were not specified, in a further 12 studies that were excluded.

RESULTS

In total, 21 of the identified 57 papers were included in the review. A further nine papers were added from the reference lists of these 21 papers and the author's personal records. A total of 30 papers form the content of this review.

Initially, we provide a summary of cognitive screening measures for MCI (Table 1), starting with the most widely used cognitive screening instruments. An examination of the usefulness of such measures (grouped in accordance with their degree of domain coverage and single or combinations of measures) follows thereafter (Table 2). It is recognised that cognitive screening measures are used for different purposes in different settings and at varying levels of specialist care. Within research and general practice contexts the emphasis is on the identification of clinically significant levels of impairment across a restricted range of cognitive abilities. At more specialist levels of care, knowledge gained via more comprehensive assessment of each of the primary domains of cognition (i.e. memory, language, visuospatial/perceptual processing, attention and executive functioning) is utilised in the differential diagnostic process. For practical reasons then, the screening instruments that form part of this review are grouped in accordance with their comprehensiveness. Those providing coverage of each of the primary domains of cognitive function are

DOI: 10.1002/gps

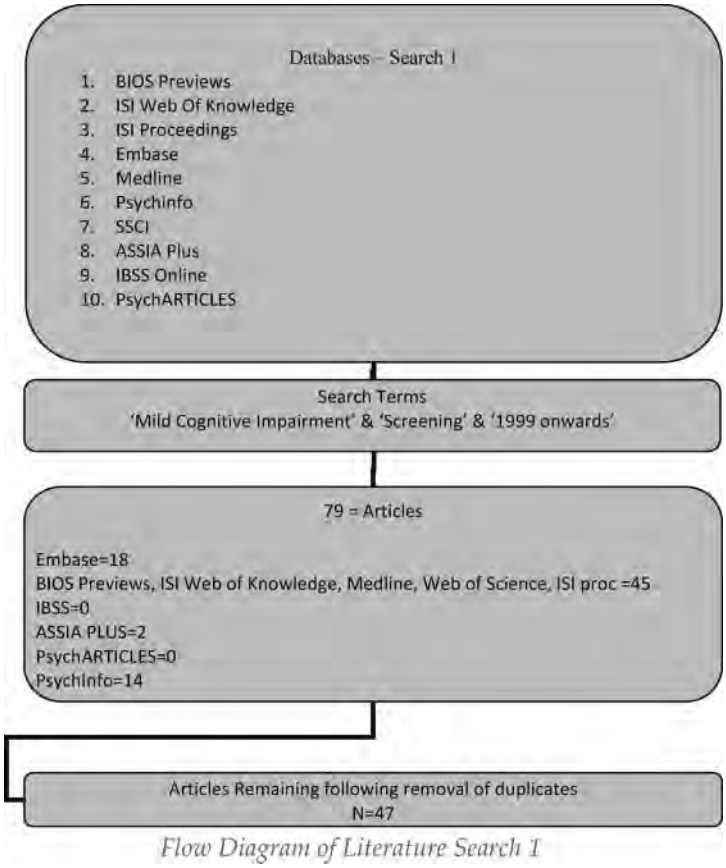


Figure 1. Flow Diagram of Literature Search 1.

referred to as 'comprehensive' and those providing partial domain coverage as 'non-comprehensive'.

Table 1 lists the range of cognitive screening measures identified, grouped in accordance with their comprehensiveness: Relatively few provide coverage of the full range of cognitive domains, as established by neuropsychological profiles in different dementias (Roth et al., 1986; Morris et al., 1989; Nasreddine et al., 2005; Mioshi et al., 2006), requiring administration times of over 10 min. The bulk of screening measures identified in this review provide an indication of clinically significant levels of cognitive impairment based on performance across a restricted range of cognitive domains, without allowing for more detailed analysis of patterns of performance (administration times 2–15 min). With the exception of the CDT, recent memory is universally assessed, although there is a good deal of variability in paradigms, ranging from registration and later recall of words, sentences, test instructions, to name and address. Recognition or cued recall items together with some measure of language or semantic memory function and attention/working memory, form part of each of the comprehensive screening measures and the lengthier (i.e. > 10 min) non-comprehensive measures, with the exception of the CAMCI (Lam et al., 2008).

In the main, copied drawings, incl. clock drawing, comprise the assessment of visuospatial

DOI: 10.1002/gps

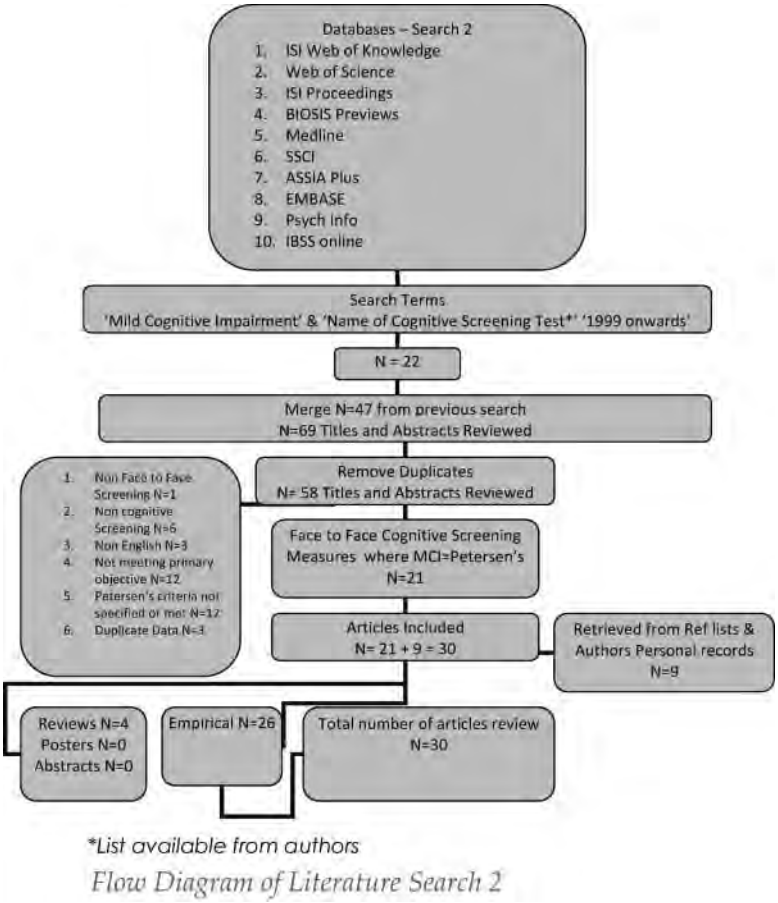


Figure 2. Flow Diagram of Literature Search 2.

and constructional skills. The ADAS-Cog (Mohs et al., 1997) provides the only assessment of praxis and the ACE-R is the only screening measures to provide an assessment of visuo-perceptual ability. None of the 15 screening measures identified includes an assessment of general reaction time.

Comprehensive cognitive screening

While high sensitivity for MCI (i.e. 84–90%) is reported for both the ACE-R and the MoCA (Nasreddine *et al*., 2005; Mioshi *et al*., 2006;), all but one of the subtests from the CERAD failed to discriminate between a small group ($n = 15$) of minimally educated MCI patients and age and education matched healthy controls (Karrasch *et al*., 2005). Adequate levels of sensitivity for MCI have also been reported for the CAMCOG subtest from the CAMDEX (Neilson *et al*., 1999). High levels of specificity are reported for both the ACE-R and MoCA in relation to normal ageing alongside moderate values (72%) for the CAMCOG (Table 2). Specificity values are not reported for the CERAD. ACE-R performance

DOI: 10.1002/gps

Table 1. Cognitive domain and item coverage for the screening measures reviewed

Cognitive Screening Measure	Orientation (number of items)	Memory (number of items)	Semantic Memory/ Language	Attention/ Calculation/ Transcoding	Reasoning/ Fluency/ Abstraction	Visuo-spatial/Visuo-constructional/ Praxis	Processing Speed/Reaction Time	Total Score	Admin. Time [min]
Comprehensive Screening Tests									
ACE-R	Temporal (5), Topographic (5)	Word registration/ recall (3), Name and address (7)	Item naming, Semantic probe questions, Command following, Irregular word reading, Sentence/Word repetition, Sentence reading/writing	Backward spelling OR Serial subtraction	Fluency: 'P' letter, 'Animal' category	Copy: Necker cube, Overlapping pentagons, Drawing: Clock	—	100	12–20
MoCA	Temporal (4), Topographic (5)	Word registration/ Recall (3)	Item naming, Sentence repetition	Short TMT B, Digit Span Forward/Backward, Tapping task, Serial subtraction	Abstraction: similarities	Copy: Cube, Drawing: Clock	—	30	10–12
CERAD	Temporal (5), Topographic (5)	Word registration/ Recall (3), Word list	Item naming, Sentence repetition, Command following, Sentence writing/reading	Backward spelling OR Serial subtraction	Fluency: 'Animal' Category	Copy: Overlapping pentagons; 2D and 3D figures	—	100	30
Camcog	Temporal (5), Topographic (5)	Word registration/ Recall (3)						105	20
Non-Comprehensive Screening Tests									
MMSE	Temporal (5), Topographic (5)	Word registration (3)/Word Recall (3)	Item Naming (2), command following (1), sentence reading (1), sentence writing (1)	Backward spelling (5)	—	Pentagon copying (1)		30	10
CDT	—	—	—	Hand and number placement	—	Clock drawing	—	Varies with scoring method	5
M@T	Temporal (5)	Sentence registration/recall (Free and cued), Word registration/ Recall (5)	General knowledge questions	—	—	—	—	50	5
SIS	Temporal (3)	Word registration/ Recall (3)	—	—	—	—	—	6	1–2
DemTect	—	Word registration/ Recall (10)	—	Number Transcoding, Digit Span backward	Fluency: 'Supermarket' Category	—	—	18	8–10

(*Continues*)

DOI: 10.1002/gps

Table 1. (Continued)

Cognitive Screening Measure	Orientation (number of items)	Memory (number of items)	Semantic Memory/ Language	Attention/ Calculation/ Transcoding	Reasoning/ Fluency/ Abstraction	Visuo-spatial/Visuo-constructional/ Praxis	Processing Speed/Reaction Time	Total Score	Admin. Time [min]
ABCS	Temporal (4), Topographic (1)	Word registration/ Recall (5)	—	—	Fluency: Animal Category	Drawing: Clock	—	5	5
STMS	Temporal (3), Topographic (3), Person (2)	Word registration/ Recall (4)	General knowledge questions	Digit Span forward, Calculation	Abstraction: similarities	Copy: cube Drawing: clock	—	38	5–7
ADAS-cog	Temporal (6), Topographic (1)	Word registration/ Recognition (10)	Spoken language: production, comprehension, word finding	—	—	Ideational praxis (well learned skills)	—	70	NR
Neurocog		4 trial word list recall	NS	NS	NS	—	—	NR	< 10
CAMCI	—	10 minute delayed 10 item word list recall		—	Animal fluency	—	—	NR	15
ECR		4 trials of semantically cued recall for 4 visual stimuli/trial, followed by delayed free and then semantically cued recall of the visual stimuli	—	—	—	—	—	NR	NR

ABCS = AB Cognitive Screen; ACE-R = Addenbrooke's Cognitive Examination Revised; ADAS-cog = Alzheimer's Disease Cooperative Study—Cognitive subscale; Cam-cog = Cambridge Cognitive Examination; CAMCI = Chinese Abbreviated Mild cognitive Impairment Test; CANS-MCI = Computer Administrated Screen for Mild Cognitive Impairment; CERAD = Consortium to Establish a registry for Alzheimer's Disease; CCSE = Cognitive Capacity Screening Examination; CMC = Combined MMSE State Examination and Cognitive Capacity Screening Examination; ECR = Enhanced Cued Recall; FCSRT = Free and Cued Selective Reminding Test; M@T = Memory Alteration Test; MCI = Mild Cognitive Impairment; MIS = Memory Impairment Screen; MoCA = Montreal Cognitive Assessment; NR = Not reported; NS = Not specified; SIS = Six Item Screen; STMS = The Short Test of Mental Status; TMT B = Trail making Test Part B.

DOI: 10.1002/gps

Table 2. Summary of study characteristics and findings for research investigating the utility of cognitive screening measures in MCI

Cognitive Screening Measure	First Author (Year)	MCI [N]	MCI Mean Age [Years] (SD)	MCI Mean Education level [Length in years] (SD)	MCI Mean MMSE Score (SD)	Comparison groups	Sample Source	Sensitivity	Specificity	Diagnostic Accuracy (AUC)	Cut-off value
Comprehensive Screening Tests											
ACE-R	Mioshi *et al.* (2006)	36	68.8 (9.0)	12.8 (3.4)	27.7 (1.5)	MCI *vs* NC	Secondary referral clinic	0.84	1.00		82/100
CAMCOG	Marcos *et al.* (2006)	82	77.6 (6.1) (CV)	84.2% <10 (CV)	25.8 (2.8) (CV)	MCI-CV *vs* MCI-NCV	Neurologist; Geriatrician specialists	0.92	0.72		79.5
			72.9 (6.8) (NCV)	77.3% <10 (NCV)	29.4 (3.5) (NCV)						
MoCA	Nasreddine *et al.* (2005)	94	75.19 (6.27)	12.28 (4.32)	27.00 (1.8)	MCI *vs* NC	University/ Community clinics	0.90	0.87		26/30
CERAD	Karrasch *et al.* (2005)	15	67.50 (9.2)	8.2 (2.1)	26.50 (2.3)	MCI *vs* NC	University medical samples	N/A	N/A		N/A
	Chandler *et al.* (2007)	60	72.80 (7.5)	14.80 (2.8)	27.50 (1.8)	A. MCI *vs* NC	Community sample	A. 0.81	A. 0.72		85.1/100
						B.MCI *vs* AD		B. 0.80	B. 0.81		
Non-comprehensive Screening Tests											
CDT	Beinhoff *et al.* (2005)	48	66.4 (7.1)	1.87* (0.94)	28.3 (1.5)	MCI *vs* NC	University Outpatient Memory Clinic	0.40		0.57	> 1
	Sager *et al.* (2006)	69	78.3 (6.7)	13.3 (3.0)	27.3 (1.9)	MCI *vs* NC	Outpatient Memory Clinics	0.2	0.88		< 8
	Ravaglia *et al.* (2005)	sMCI = 18	76.3 (7.4)	6.3 (3.4)	25.4 (2.6)	sMCI	University Outpatient Memory Clinic	0.26	0.85		6 or below
		aMCI = 38	76.5 (7.1)	8.0 (4.4)	25.7 (3.1)	aMCI		0.06			
		mMCI = 57	78.6 (8.0)	5.7 (3.3)	24.3 (2.8)	mMCI *vs* NC		0.40			
	Yamamoto *et al.* (2004)	aMCI = 10	74.6 (7.2)	MCI all 11.5(3.7)	MCI all 27.2(2.1)	MCI and dementia *vs* NC	Outpatient Memory Clinic	0.75	0.76		7
		sMCI = 10	74.0 (7.2)								
		mMCI = 28	75.0 (5.6)								
DemTect	Kalbe *et al.* (2004)	97	72.1 (9.0)	10.0 (1.2)	26.9 (1.5)	MCI v. NC		0.80	0.92		13
CAMCI	Lam *et al.* (2008)	S1 = 182	78.9 (7.0)	2.8 (4.2)	23.9 (3.2)	NC *vs* MCI and Mild dementia	Community dwelling volunteers	NR	NR	S1 = 0.91	15/16
		S2 = 162	72.8 (6.5)	3.1 (3.4)	24.5 (2.7)					S2 = 0.98	

(*Continues*)

DOI: 10.1002/gps

Table 2. (Continued)

Cognitive Screening Measure	First Author (Year)	MCI [N]	MCI Mean Age [Years] (SD)	MCI Mean Education level [Length in years] (SD)	MCI Mean MMSE Score (SD)	Comparison groups	Sample Source	Sensitivity	Specificity	Diagnostic Accuracy (AUC)	Cut-off value
ECR	Saka *et al.* (2006b)	18	69.4 (8.3)	8.4 (5.0)	26.6 (1.7)	MCI *vs* NC	Outpatient Memory Clinic	0.56	0.79	0.69	9 -3rd free recall trial
M@T	Rami *et al.* (2007a); Rami *et al.* (2007b)	50	76.6 (6.6)	8.4 (5.2)	25.1 (2.4)	A. MCI *vs* NC		A. 0.96	A. 0.79		A. 37
						B. MCI *vs* AD		B. 0.87	B. 0.82		
SIS	Callahan *et al.* (2002)	N/A	N/A	N/A	N/A	A. CI *vs* No CI	1. Community	A1. 50.4	A1. 97.4		All > 3 errors
						B. Dem *vs* No Dem.	2. Clinic	A2. 74.2	A2. 96.0		
								B1. 88.7 B2 80.6	B1. 88.0 B2. 90.9		
Mini-Cog	Borson *et al.* (2005)	71	N/A	N/A	N/A	MCI v NC		0.58	N/A		< 3
ABCS	Molloy *et al.* (2005)	124	77.2	12.1	27.2	MCI *vs* NC		0.80	0.5		N/A
			N/A	N/A	N/A						
	Standish *et al.* (2007)	166	N/A	N/A	27.1	MCI *vs* NC	Speciality geriatric clinics	N/A	N/A		N/A
					N/A						
STMS	Tang-Wai *et al.* (2003)	129	79.5 (7.2)	13.3 (3.2)	26.3 (2.2)	MCI/AD *vs* NC	Speciality geriatric clinic	N/A			N/A
ADAS-cog	Pyo *et al.* (2006)	135	70.4 (10.0)	12.8 (3.1)	26.4 (2.6)	MCI *vs* NC	Community dwelling volunteers	0.73	0.89		6 (Total Score)
	Grundman (2004)	769	72.9 (7.3)	14.7 (3.1)	27.3 (1.9)	MCI *vs* NC	Community dwelling volunteers	N/A	N/A		N/A
	Fleisher *et al.* (2007)	539	74.9 (6.6) (P)	14.5 (3.1) (P)	N/A	MCI-P *vs* MCI NP	Community dwelling volunteers	N/A	N/A		N/A
			71.5 (7.4) (NP)	14.97 (2.8) (NP)							

(*Continues*)

DOI: 10.1002/gps

Table 2. (Continued)

Cognitive Screening Measure	First Author (Year)	MCI [N]	MCI Mean Age [Years] (SD)	MCI Mean Education level [Length in years] (SD)	MCI Mean MMSE Score (SD)	Comparison groups	Sample Source	Sensitivity	Specificity	Diagnostic Accuracy (AUC)	Cut-off value
Animal Fluency	Sager *et al.* (2006)	69	78.3 (6.7)	13.3 (3.0)	27.3 (1.9)	MCI *vs* NC	Outpatient Memory Clinic	0.54	0.88	NC = 0.76	< 14
										MCI = 0.78	< 17
Combinations of Screening Tests											
MMSE and CDT	Ravaglia *et al.* (2005)	sMCI = 18	76.3 (7.4)	6.3 (3.4)	25.4 (2.6)	sMCI	University Outpatient Memory Clinic	0.45	0.69		MMSE < 24 or CDT = or < 5
		aMCI = 38	76.5 (7.1)	8.0 (4.4)	25.7 (3.1)	aMCI		0.56	0.69		
		mMCI = 57	78.6 (8.0)	5.7 (3.3)	24.3 (2.8)	mMCI *vs* NC		0.75	0.69		
MIS,LST,VF or CDT	Beinhoff *et al.* (2005)	48	66.4 (7.1)	1.87* (0.94)	28.3 (1.5)	MCI *vs* NC	University Outpatient Memory Clinic	0.83	0.74		< 5,< 3, < 20 or > 2
MIS or VF								0.75	0.84		
VF discrepancy scores	Lonie *et al.* (2008a)	47	73.9 (6.37)	NR	28.3 (1.48)	MCI *vs* NC	Outpatient Memory Clinic	NR	NR	0.84	NR
						MCI *vs* Dep				0.75	

ABCS = AB Cognitive Screen; ACE-R = Addenbrookes Cognitive Examination-Revised; AD = Alzheimer's Disease; ADAS-cog = Alzheimer's Disease Cooperative Study—Cognitive subscale; aMCI = amnestic MCI; Camcog = Cambridge Cognitive Examination;CDT = Clock Drawing Test; CERAD = Consortium to Establish a registry for Alzheimer's Disease; CI = All participants with cognitive impairment; CV = MCI converter to dementia; Dem. = Dementia; Dep = Depression; .* Education 1 = up to 10 years of education, 2 = more than 10 and up to 13 years of education, 3 = University degree; ES = English Speaking; FU = follow-up; LST = Letter Sorting Test; M@T = Memory Alteration Test; MCI = Mild Cognitive Impairment; MIS = Memory Impairment Screen; MMSE = Mini Mental Cognitive Examination; mMCI = multiple domain MCI; MoCA = Montreal Cognitive Assessment; N/A = Not assessed; NC = Healthy Control Participant; NCV = MCI non-converter to dementia; NP = Non Progressive; P = Progressive; S1 = sample 1; SD = Standard Deviation; sMCI = single non-memory domain MCI; SIS = Six Item Screen; SS = Spanish speaking; STMS = The Short Test of Mental Status; VF = verbal fluency (animals).

DOI: 10.1002/gps

data are available only for a small number ($n = 36$) of well educated, relatively young MCI sufferers of above average pre-morbid intellectual functioning (Mioshi *et al*., 2006). The applicability of the normative data to older and less well educated MCI sufferers of average pre-morbid IQ remains uncertain.

The CAMCOG is the only comprehensive cognitive screening measure for which predictive validity values are reported, although the level of specificity for conversion to dementia would appear to be too low to support its use for prognostic purposes (Neilson *et al*., 1999). None of the measures report specificity values for MCI in relation to depression or any other neurological or psychiatric diseases of a non-progressive nature.

Moderate to high 1-month test–retest reliability values (70–92%) are reported for MoCA (Nasreddine *et al*., 2005). Despite the need for longitudinal follow-up in this specific patient group, reliability data are not reported for clinically relevant intervals (i.e. 6–12 months) or for any other of the more comprehensive cognitive screening measures. The absence of this information complicates the interpretation of score changes that are so important in determining prognosis in this patient group.

Two comprehensive cognitive screening measures, the ACE-R (Mioshi *et al*., 2006) and the DemTect (Kalbe *et al*., 2004) have been validated for use within a range of non-AD dementias in conjunction with MCI. This is particularly useful in secondary and tertiary care where cognitive screening routinely forms part of the differential diagnostic process.

Without exception, patient samples for studies investigating the more comprehensive screening measures were drawn from specialist university or hospital clinics. Their utility in screening for MCI within community or primary care settings is therefore unknown. Cut-off values with corresponding levels of sensitivity and specificity to assist in differentiating MCI from normal aging are provided for all but the CERAD. Demographic adjustment equations are provided for the MoCA (Nasreddine *et al*., 2005), whilst cut-off values for the CAMCOG and the ACE-R relate to raw scores, facilitating ease of use in busy clinic settings.

Non-comprehensive cognitive screening

Table 2 also summarises the relevant characteristics of the cognitive screening measures with incomplete domain coverage. Of these, the MMSE (Folstein *et al*., 1975) is the most commonly used in clinical practice (Shulman *et al*., 2006). Its ability to differentiate MCI sufferers from healthy elderly controls has not been established consistently (Saka *et al*., 2006; Slavin *et al*., 2007). Furthermore, significant differences, where they do exist, range in magnitude from less than 1 scale point (Ravaglia *et al*., 2005) to a maximum of just under 2 points (Slavin *et al*., 2007). As the majority of MCI patients score above the commonly used MMSE cut-offs of 24 and 26, there is a considerable overlap in the scores of patients with MCI and age matched healthy controls. The sensitivity of the MMSE to MCI is therefore low with few exceptions (Callahan *et al*., 2002), ranging between 1% (Sager *et al*., 2006) and 49% (Ravaglia *et al*., 2005). Variability in reported sensitivity rates is mainly related to different MMSE cut-off values and comparison groups, such as a combined group of MCI and early dementia sufferers (Lam *et al*., 2008) or differing proportions of 'pre-clinical dementia' patients. Correspondingly, MMSE specificity for MCI compared with healthy volunteers is generally high, ranging from 85.5–100% (Sager *et al*., 2006). Scores below the commonly suggested MMSE cut-off of 24 or 26 will signify the presence of a dementing illness (and indeed MCI), while scores above this level cannot be assumed to reflect the absence of either condition. Tang-Wai *et al*. (2003) found no significant difference in the baseline MMSE scores of well individuals who later developed MCI and those who didn't. Direct comparison of the MMSE with alternative screening measures indicates that the MMSE is less effective at discriminating between MCI patients and healthy age matched controls (Tang-Wai *et al*., 2003; Kalbe *et al*., 2004; Nasreddine *et al*., 2005; Rami *et al*., 2007; Standish *et al*., 2007), between persons with progressive and non-progressive forms of cognitive impairment (Tang-Wai *et al*., 2003; Rami *et al*., 2007) and is more prone to the influences of age and education (Molloy *et al*., 2005) than a range of other cognitive screening measures.

Levels of sensitivity and specificity comparable to the comprehensive screening measures are reported for some of the other non-comprehensive screens, i.e. CDT, DemTect, Memory Alternation Test – M@T (Rami *et al*., 2007), and ADAS-Cog. Sensitivity and specificity values are not reported for the Short Test of Mental Status (STMS; Tang-Wai *et al*., 2003) or the CAMCI (Lam *et al*., 2008). For a further two, CDT and the AB Cognitive Screen (ABCS; Molloy *et al*., 2005), one or more of these values are too low to support their use in MCI screening. The single set of favourable sensitivity and specificity values reported for the CDT (Yamamoto *et al*., 2004) together with AUCs reported in association with the CAMCI (Lam *et al*., 2008) is

DOI: 10.1002/gps

likely to reflect the contamination of the MCI group with patients in the early stages of dementia, i.e. in each of these studies a combined group of MCI and early dementia sufferers was compared with an aged matched normal control group. The Six Item Screener (SIS; Callahan *et al*., 2002) would also appear to be of limited usefulness in detecting MCI amongst the elderly in the community. Considerably higher sensitivity and specificity are reported for the SIS in a specialist clinic setting, however the restricted range of items (i.e. orientation and memory questions only), would possibly prove inadequate for use at secondary and tertiary levels of assessment.

Administration times are on the whole shorter, i.e. less than 10 min, with a majority (CDT, M@T, SIS, ABCS and STMS) reporting minimum administration times of 5 min or less, simple scoring and administration methods without necessity for training, making these potentially well suited to use in general practice. Predictive validity data are provided for two of the measures, based on longitudinal follow-up periods of between 3–5.5 years. Correct classification rates of around 70% are reported for both the STMS (Tang-Wai *et al*., 2003) and ADAS-Cog (Fleisher *et al*., 2007).

Whilst issues relating to the differential diagnosis of MCI may be of less relevance in primary care where an emphasis is placed on the 'need to refer', the majority of non-comprehensive cognitive screening measures have also been validated for use as screening tools in dementia, for the most part Alzheimer's Disease, and in a few cases (i.e. DemTect and ABCS) for non-AD dementias.

Reliability data are absent with exception of the high inter-rater ($r = 0.99$) and 1 month test–retest reliabilities ($r = 0.92$) reported for the DemTect (Kalbe *et al*., 2004) and MoCA (Nasreddine *et al*., 2005a) respectively. Practice effects in the order of less than one point over 6 and 12-month intervals are also reported for the DemTect (Kalbe *et al*., 2004) that would benefit from replication in larger more uniform samples. It is impossible to know what size of change in test scores would represent a clinically significant level of cognitive deterioration, i.e. how much variability in a patient's performance is attributable to inconsistencies in test administration or scoring procedures between clinicians.

Cut-off values for the detection of MCI (across one or more combinations of sensitivity and specificity) are reported for the bulk of the less comprehensive screening measures with the exception of the STMS and ABCS. Further breakdown of cut-off scores for MCI in accordance with base rates (i.e. the frequency with which the condition might be expected to occur within the relevant setting), is universally absent.

The age range of MCI samples appears to be broadly reflective of that reported in the wider literature (Alladi *et al*., 2006; Lonie *et al*., 2008b) and relatively consistent across MCI screening studies. Mean MMSE scores for the MCI patient groups range from 23 to of 29. The range of mean MMSE scores indicates that even within Petersen's MCI criteria there remains scope for significant variability in terms of cognitive ability, with the result of influencing reported levels of sensitivity. Wide variation in education levels further complicates the interpretation of MMSE scores. A mean MMSE score of 24 in highly educated patients will signify a more advanced stage of disease than in patients with fewer years of education. Close consideration should therefore be given to the mean education and MMSE scores of MCI patient groups when comparing the performance of screening measures across studies.

With the exception of the DemTect (Kalbe *et al*., 2004), none of the most recent versions of the brief screening measures have been validated for use across more than one language or culture. The SIS represents the only cognitive screening measure whose ability to detect MCI has been studied across multiple settings, i.e. community and primary care (Callahan *et al*., 2002).

Combinations of screening tests

Several studies have examined the possibility that administration and interpretation of a patient's performance on a combination of two or more cognitive screening instruments may yield superior classification accuracy than a single measure. Ravaglia *et al*. (2005) compared the single and combined ability of the MMSE and CDT to differentiate between healthy elderly, Alzheimer's Disease patient's and three MCI subtypes. Whereas the sensitivity of both the clock drawing test and MMSE to MCI was reportedly universally poor ($< 50\%$), when performance on both screening instruments was evaluated, an improvement in sensitivity to the multi-domain MCI subtype was noted (75%). Predictive validity for the remaining MCI subtypes remained low, however, and the authors caution against the use of this combination of instruments in routine MCI screening.

Beinhoff *et al*. (2005) report combinations of moderate to high sensitivity and specificity using a combination of the Memory Impairment Screen (comprising the delayed free and cued recall of four items) and an animal fluency task. By administering a letter sorting task (requiring the patient to spell a five

DOI: 10.1002/gps

digit word, forwards, backwards and then in alphabetical order), in conjunction with the above two measures, sensitivity to MCI rose to a clinically acceptable level (i.e. 83%) at the expense of a decline in specificity (i.e. from 84–74%).

Finally, several studies have examined the ability of performance differentials between lexical (letter fluency) and semantic (category) fluency to differentiate MCI sufferers from healthy elderly controls. Murphy *et al*. (2006) reported that difference scores between generation of words according to the category of 'animals' or the letter 'F' distinguished MCI patients from controls with moderate effect size. Using a semantic (category) fluency advantage of less than one standard deviation or more than the mean for the control group as a cut off however, sensitivity to MCI was too low (i.e. 30%) for screening purposes. Moderate to high levels of diagnostic accuracy in differentiating aMCI sufferers from healthy elderly controls (area under the curve (AUC) = 0.837) and elderly depressed patients (AUC = 0.71) using fluency discrepancy scores ('animals' vs. 'P' words) have also been reported by Lonie *et al*. (2008a). The AUCs reported in this study were as high as those obtained using a standardised measure of delayed recall. Whilst all but two control group patients generated a greater number of animals than 'P' words across the one minute time span provided, a good deal of directional variability was present in the MCI and AD patient groups. Without knowledge of the fate of the MCI subjects comprising these studies, it remains difficult to ascertain the potential clinical utility of fluency discrepancy scores in the cognitive screening assessment of MCI.

CONCLUSION

The dearth of validated MCI screening measures is problematic in view of the high numbers of community dwelling cognitively impaired but not demented older adults (Busse *et al*., 2003b; Lopez *et al*., 2003) and the significant proportions of MCI presenting to secondary and tertiary specialist clinics (Alladi *et al*., 2006; Lonie *et al*., 2008b).

At the most basic level, a cognitive screening instrument should provide an indication of the likely presence of clinically meaningful cognitive impairment, and hence the need to refer on to more specialist services. Of the 15 screening measures identified, sensitivity and specificity values for MCI as defined by Petersen *et al*. (1999), compared with normal elderly controls, were adequate (i.e. $\geq$ 79%) in four cases. Each of these four cognitive screening measures (ACE-R, DemTect, MoCA, and M@T) requires average administration times of 15 min or less. In two instruments (M@T and ACE-R) cut-off values are reported without the need for transformation of raw scores. Adequate sensitivity to a wider range of early dementia presentations is reported for the ACE-R only and specificity values for MCI in relation to depression are universally absent. This is surprising in view of the frequency with which MCI and depression are known to coexist (Solfrizzi *et al*., 2007) and the well documented difficulties clinicians face in attempting to tease apart their respective contributions to cognitive impairment (O'Carroll *et al*., 1994; Lezak 1995).

In more specialist clinical settings, information gathered through the administration of cognitive screening measures forms an integral part of the clinical evaluation and differential diagnosis. Four of the 15 cognitive screening measures identified in this review cover domains comprehensively, thus facilitating this function. For each of these screening measures adequate sensitivity to MCI and early AD is reported. However, without exception, there is an absence of reliability data for elderly persons who are not cognitively impaired, and of the predictive validity of cut-off scores. Both factors limit the interpretation of score changes and the ability to supply prognostic information. The absence of subtests assessing general reaction time within cognitive screening measures is especially remarkable, as a reduction in processing speed appears to predate clinical onset of Alzheimer's disease (Backman *et al*., 2005; Rami *et al*., 2007) and is a good predictor of which MCI sufferers will develop dementia (Tabert *et al*., 2006).

Future research should focus on establishing a wider range of psychometric test properties (i.e. reliability and predictive validity) for those cognitive screening measures where adequate sensitivity and specificity exist. This would equip the clinician with the ability to interpret the meaning of score changes and provide some opportunity of likely prognosis. It would also be of interest to compare the screening capacity of further combinations of brief cognitive screening measures or formal neuropsychological measures to the screening measures described above, as it may be that some of the equally brief neuropsychological measures perform as well as, if not better than cognitive screening instruments (Lonie *et al*., 2008a). Furthermore, the present review findings relate to the performance of cognitive screening instruments in the assessment of the single and multi-domain amnestic MCI subtypes. As the relevance of identifying specific MCI subtypes emerges, it will be important to establish the screening

DOI: 10.1002/gps

power of the above cognitive measures in MCI's non-amnestic forms.

None of the identified cognitive screening measures wholly fulfil all of the criteria we have identified as being important in MCI screening. Whilst the sensitivity of the instruments covering all cognitive domains is high, we miss data about sensitivity for early atypical dementia presentations, specificity compared with psychiatric and non-progressive neurological conditions, cross-cultural usage, reliability and predictive validity. Consequently, whilst several cognitive screening instruments afford the clinician the ability to detect MCI, early AD, and in some cases non-AD dementia, they cannot currently be used to make reliable inferences about the course and eventual outcome of MCI.

CONFLICT OF INTEREST

None known.

ACKNOWLEDGEMENTS

This study was supported by the Gordon Small Charitable Trust and the Medical Research Council (UK), and the European Commission Network of Excellence ''Diagnostic Molecular Imaging'' (FP6-LIFESCIHEALTH Project Reference: 512146).

REFERENCES

Albert MS, Blacker D. 2006. Mild cognitive impairment and dementia. *Ann Rev Clin Psychol* **2**: 379–388.

Alladi S, Arnold R, Mitchell J, *et al*. 2006. Mild cognitive impairment: applicability of research criteria in a memory clinic and characterization of cognitive profile. *Psychol Med* **36**: 507–513.

Backman L, Jones S, Berger A, *et al*. 2005. Cognitive impairment in preclinical Alzheimer's disease: a meta-analysis. *Neuropsychology* **19**: 520–531.

Beinhoff U, Hilbert V, Bittner D, *et al*. 2005. Screening for cognitive impairment: a triage for outpatient care. *Dement Geriatr Cogn Disord* **20**: 278–285.

Bennett DA, Schneider JA, Bienias JL, *et al*. 2005. Mild cognitive impairment is related to Alzheimer's disease pathology and cerebral infarctions. *Neurology* **64**: 834–841.

Borson S, Scanlan JM, Watanabe J, *et al*. 2005. Simplifying detection of cognitive impairment: comparison of the mini-cog and mini-mental state examination in a multiethnic sample. *J Am Geriatr Soc* **53**: 871–874.

Busse A, Angermeyer MC, Riedel-Heller SG. 2006. Progression of mild cognitive impairment to dementia: a challenge to current thinking. *Br J Psychiatry* **189**: 399–404.

Busse A, Bischkipf J, Riedel-Heller SG, *et al*. 2003a. Subclassifications for mild cognitive impairment: prevalence and predicitive validity. *Psychol Med* **33**: 1029–1038.

Busse A, Bischkopf J, Riedel-Heller SG, *et al*. 2003b. Mild cognitive impairment: prevalence and incidence according to different diagnostic criteria: Results of the Leipzig Longitudinal Study of the Aged (LEILA75+). *Br J Psychiatry* **182**: 449–454.

Callahan CM, Unverzagt FW, Hui SL, *et al*. 2002. Six-Item screener to identify cognitive impairment among potential subjects for clinical research. *Medical Care* **40**: 771–781.

Chandler MJ, Lacritz LH, Hynan LS, *et al*. 2007. A total score for the CERAD neuropsychological battery. *Neurology* **65**: 102–106.

Cullen B, O'Neill B, Evans JJ, *et al*. 2007. A review of screening tests for cognitive impairment. *J Neurol Neurosurg Psychiatry* **78**: 790–799.

Fleisher AS, Sowell BB, Taylor C, *et al*. 2007. Clinical predictors of progression to Alzheimer disease in amnestic mild cognitive impairment. *Neurology* **68**: 1588–1595.

Folstein M, Folstein M, McHugh P. 1975. 'Mini-Mental State'. A practical method for grading the cognitive state of patients for the clinician. *J Psychiat Res* **12**: 189–198.

Frisoni GB, Fratiglioni L, Fastbom J, *et al*. 2000. Mild Cognitive Impairment in the Population and Physical Health: Data on 1,435 Individuals Aged 75 to 95. *J Gerontol A Biol Sci Med Sci* **55**: M322–M328.

Galvin JE, Roe CM, Morris JC. 2007. Evaluation of cognitive impairment in older adults. *Arch Neurol* **64**: 718–724.

Giaquinto S, Parnetti L. 2006. Early detection of dementia in clinical practice. *Mech Ageing Dev* **127**: 123–128.

Grundman M, Petersen RC, Ferris SH, *et al*. 2004. Mild Cognitive Impairment Can Be Distinguished From Alzheimer Disease and Normal Aging for Clinical Trials. *Arch Neurol* **61**: 59–66.

Hirao K, Ohnishi T, Hirata Y, *et al*. 2005. The prediction of rapid conversion to Alzheimer's disease in mild cognitive impairment using regional cerebral blood flow SPECT. *NeuroImage* **28**: 1014–1021.

Jessen F, Feyen L, Freymann K, *et al*. 2006. Volume reduction of the entorhinal cortex in subjective memory impairment. *Neurobiol Aging* **27**: 1751–1756.

Kalbe E, Kessler J, Calabrese P, *et al*. 2004. DemTect: A new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia. *Int J Geriatr Psychiatry* **19**: 136–143.

Karrasch M, Sinerva E, Gronholm P, *et al*. 2005. CERAD test performances in amnestic mild cognitive impairment and Alzheimer's disease. *Acta Neurol Scand* **111**: 172–179.

Lam LCW, Tam CWC, Lui VWC, *et al*. 2008. Screening of mild cognitive impairment in Chinese older adults – A multistage validation of the Chinese abbreviated mild cognitive impairment test. *Neuroepidemiology* **30**: 6–12.

Lehrner J, Gufler R, Guttmann G, *et al*. 2005. Annual conversion to Alzheimer disease among patients with memory complaints attending an outpatient memory clinic: the influence of amnestic mild cognitive impairment and the predictive value of neuropsychological testing. *Wiener Klin. Wochenschr.* **117**: 629–635.

Lezak MD. 1995. *Neuropsychological Assessment*. Oxford University Press: London.

Lines CR, McCarroll KA, Lipton RB, *et al*. 2003. Telephone screening for amnestic mild cognitive impairment. *Neurology* **60**: 261–266.

Lonie JA, Hermann LL, Tierney KM, *et al*. 2008a. Lexical and semantic fluency discrepancy scores in aMCI and early Alzheimer's disease. *J Neuropsychol* in press.

Lonie JA, Herrmann LL, Donaghey CL, *et al*. 2008b. Mild cognitive impairment referral patterns and cognitive profile: Some implications for clinical practice. *Br J Psychiatry* **192**: 59–64.

Lopez OL, Jagust WJ, DeKosky ST, *et al*. 2003. Prevalence and classification of mild cognitive impairment in the Cardio-vascular Health Study Cognition Study. *Arch Neurol* **60**: 1385–1389.

Marcos A, Gil P, Barabash A, *et al*. 2006. Neuropsychological markers of progression from mild cognitive impairment to Alzheimer's disease. *Am J Alzheimers Dis Other Demen.* **21**: 189–196.

Mioshi E, Dawson K, Mitchell J, *et al*. 2006. The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *Int J Geriatr Psychiatry* **21**: 1078–1085.

Mitchell T, Woodward M, Hirose Y. 2008. A survey of attitudes of clinicians towards the diagnosis and treatment of mild cognitive impairment in Australia and New Zealand. *Int Psychogeriatr* **20**: 77–85.

Mohs RC, Knopman D, Petersen RC, *et al*. 1997. Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's Disease Assessment Scale (ADAS) that broaden its scope: the Alzheimer's Disease Cooperative Study. *Alzheimer Dis Assoc Disord.* **11** (Suppl 2): 13S–21S.

Molloy DW, Standish TIM, Lewis DL. 2005. Screening for mild cognitive impairment: Comparing the SMMSE and the ABCS. *Can J Psychiatry* **50**: 52–58.

Morris JC, Storandt M, Miller JP, *et al*. 2001. Mild cognitive impairment represents early-stage Alzheimer's disease. *Arch Neurol* **58**: 397–405.

Morris JC, Heyman A, Mohs RC, *et al*. 1989. The consortium to establish a registry for Alzheimer's disease (CERAD), part I: clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* **39**: 1159–1165.

Murphy KJ, Rich JB, Troyer AK. 2006. Verbal fluency patterns in amnestic mild cognitive impairment are characteristic of Alzheimer's type dementia. *J Int Neuropsychol Soc* **12**: 570–574.

Nasreddine ZS, Phillips NA, Bedirian V, *et al*. 2005. The Montreal Cognitive Assessment, MOCA: A brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* **54**: 695–699.

Neilson H, Lolk A, Andersen K, *et al*. 1999. Characteristics of elderly who develop Alzheimer's disease in the next 2 years—a neuropsychological study using CAMCOG. The Odense Study. *Intl J Geriatr Psychiatry* **14**: 957–963.

O'Carroll RE, Curran SM, Ross M, *et al*. 1994. The differentiation of major depression from dementia of the Alzheimer type using within-subject neuropsychological discrepancy analysis. *Br J Clin Psychol* **33**: 23–32.

Palmer K, Berger AK, Monastero R, *et al*. 2007. Predictors of progression from mild cognitive impairment ot Alzheimer disease. *Neurology* **68**: 1596–1602.

Petersen RC. 2004. Mild Cognitive Impairment as a diagnostic entity. *J Int Med* **256**: 183–194.

Petersen RC, O'Brien J. 2006. Mild cognitive impairment should be considered for DSM-V. *J Geriatr Psychiatry Neurol* **19**: 147–154.

Petersen RC, Smith GE, Waring SC, *et al*. 1999. Mild Cognitive Impairment: Clinical characterization and outcome. *Arch Neurol* **56**: 303–308.

Petersen RC, Stevens JC, Ganguli M, *et al*. 2001. Practice parameter: early detection of dementia: mild cognitive impairment (an evidence-based review): Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* **56**: 1133–1142.

Pyo G, Elble RJ, Ala T, *et al*. 2006. The characteristics of patients with uncertain/mild cognitive impairment on the alzheimer's disease assessment scale-cognitive scubscale. *Alzheimer Dis Assoc Disord.* **20**: 16–22.

Rami L, Gomez-Anson B, Sanchez-Valle R, *et al*. 2007a. Longitudinal study of amnesic patients at high risk for Alzheimer's disease: clinical, neuropsychological and magnetic resonance spectroscopy features. *Dement Geriatr Cogn Disord* **24**: 402–410.

Rami L, Molinuevo JL, Sanchez-Valle R, *et al*. 2007b. Screening for amnestic mild cognitive impairment and early Alzheimer's disease with MT (Memory Alteration Test) in the primary care population. *Int J Geriatr Psychiatry* **22**: 294–304.

Ravaglia G, Forti P, Maioli F, *et al*. 2005. Screening for mild cognitive impairment in elderly ambulatory patients with cognitive complaints. *Aging Clin Exp Res* **17**: 374–379.

Rose SE, McMahon KL, Janke AL, *et al*. 2006. Diffusion indices on magnetic resonance imaging and neuropsychological performance in amnestic mild cognitive impairment. *J Neurol Neurosurg Psychiatry* **77**: 1122–1128.

Roth M, Tym E, Moutjoy CQ, *et al*. 1986. CAMDEX. A standardised instrument for the diagnosis of mental disorder in the eldery with special reference to the early detection of dementia. *Br J Psychiatry* **149**: 698–709.

Sager MA, Hermann BP, La Rue A, *et al*. 2006. Screening for dementia in community-based memory clinics. *WMJ* **105**: 25–29.

Saka E, Mihci E, Topcuoglu MA, *et al*. 2006a. Enhanced cued recall has a high utility as a screening test in the diagnosis of Alzheimer's disease and mild cognitive impairment in Turkish people. *Arch Clin Neuropsychol* **21**: 745–751.

Saykin AJ, Wishart HA, Rabin LA, *et al*. 2006. Older adults with cognitive complaints show brain atrophy similar to that of amnestic MCI. *Neurology* **67**: 834–842.

Shulman KI, Herrmann N, Brodaty H, *et al*. 2006. Survey of brief cognitive sceening instruments. *Int Psychogeriatr* **18**: 281–294.

Slavin MJ, Sandstrom CK, Tran TTT, *et al*. 2007. Hippocampal volume and the Mini-Mental State Examination in the diagnosis of amnestic mild cognitive impairment. *Am J Roentgenol* **188**: 1404–1410.

Solfrizzi V, D'Introno A, Colacicco AM, *et al*. 2007. Incident occurence of depressive symptoms among patients with mild cognitive impairment—the Italian longitudinal study on aging. *Dement Geriatr Cogn Disord* **24**: 55–64.

Standish TIM, Molloy DW, Cunje A, *et al*. 2007. Do the ABCS 135 short cognitive screen and its subtests discriminate between normal cognition, mild cognitive impairment and dementia? *Int J Geriatr Psychiatry* **22**: 189–194.

Storandt M, Grant EA, Miller JP, *et al*. 2006. Longitudinal course and neuropathologic outcomes in original vs revised MCI and in pre-MCI. *Neurology* **67**: 467–473.

Tabert MH, Many JJ, Liu X, *et al*. 2006. Neuropsychological prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. *Arch Gen Psychiatry* **63**: 916–924.

Tang-Wai DF, Knopman DS, Geda YE, *et al*. 2003. Comparison of the Short Test of Mental Status and the Mini-Mental State Examination in Mild Cognitive Impairment. *Arch Neurol* **60**: 1777–1781.

van Uffelen JGZ, Chin A, Paw MJM, *et al*. 2007. Detection of memory impairment in the general population: screening by questionnaire and telephone compared to subsequent face-to-face assessment. *Int J Geriatr Psychiatry* **22**: 203–210.

Winblad B, Palmer K, Kivipelto M, *et al*. 2004. Mild cognitive impairment—beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *J Intern Med* **256**: 240–246.

Yamamoto S, Mogi N, Umegaki H, *et al*. 2004. The clock drawing test as a valid screening method for mild cognitive impairment. *Dement Geriatr Cogn Disord* **18**: 172–117.

Journal of Neuropsychology (2009), 3, 79–92
© 2009 The British Psychological Society

The British Psychological Society

www.bpsjournals.co.uk

Lexical and semantic fluency discrepancy scores in aMCI and early Alzheimer's disease

Jane A. Lonie[1], Lucie L. Herrmann[2], Kevin M. Tierney[1], Claire Donaghey[1], Ronan O'Carroll[3], Andrew Lee[4] and Klaus P. Ebmeier[1,2]*

[1]Division of Psychiatry, University of Edinburgh, Edinburgh, UK
[2]Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford, UK
[3]Department of Psychology, University of Stirling, Stirling, UK
[4]Jardine Clinic, Royal Edinburgh Hospital, Edinburgh, UK

Episodic memory is compromised in amnestic mild cognitive impairment (aMCI), but lesser deficits in other cognitive domains are also commonly observed and may be helpful in identifying this group. The relative difference in performance on lexical and semantic fluency tasks may be a sensitive and specific measure in aMCI and early Alzheimer's disease (AD). We compared four groups of participants, 35 early AD, 47 aMCI, 24 healthy controls, and 18 depressive out-patient controls, on semantic and lexical fluency as well as other neuropsychological tests. Early AD and aMCI patients showed a distinct pattern of semantic impairment in the two fluency measures compared with the healthy and depressive controls. The findings implicate early failure of the semantic memory system in aMCI and AD and suggest that consideration of the discrepancy in performance on semantic and lexical fluency measures may help in the early identification of AD.

A recent meta-analysis demonstrated that Alzheimer's disease (AD) patients are significantly more impaired on measures of semantic than lexical fluency (Henry, Crawford, & Phillips, 2004). Semantic memory impairment in AD is specific in that it is not associated with more generalized deficits in verbal intelligence or psychomotor speed. This pattern of impairment in verbal fluency measures is qualitatively distinct from the usual finding of superior semantic fluency in healthy controls (Spreen & Strauss, 1998). As the category task is thought to rely more heavily on access to representations of semantic concepts than the letter task, the pattern of findings in AD is presumed to reflect degradation in the structure, content, or activation of the semantic memory system (Auriacombe *et al.*, 2006; Jefferies & Lambon Ralph, 2006; Jones, Laukka, & Backman, 2006).

*Correspondence should be addressed to Professor Klaus P. Ebmeier, Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford OX3 7JX, UK (e-mail: klaus.ebmeier@psych.ox.ac.uk).

DOI:10.1348/174866408X289935

Patients in the preclinical stages of AD exhibit a semantic fluency deficit, at a time when lexical fluency performance remains intact (Auriacombe *et al.*, 2006; Beatty, Salmon, Troster, & Tivis, 2002; Swainson *et al.*, 2001). Similarly, patients who fulfil criteria for amnestic mild cognitive impairment (aMCI; Grundman *et al.*, 2004; Petersen *et al.*, 1999) generate fewer words from a specified category than do age-matched controls. In contrast, they perform at normal levels on lexical fluency tasks (Alladi, Arnold, Mitchell, Nestor, & Hodges, 2006; Dudas, Clague, Thompson, Graham, & Hodges, 2005; Lonie, Herrmann, Donaghey, & Ebmeier, 2008; Murphy, Rich, & Troyer, 2006).

A pattern of worse semantic than lexical fluency has also been reported in patients with depression (Christensen, Griffiths, Mackinnon, & Jacomb, 1997; Zakzanis, Leach, & Kaplan, 1998), although a more recent review suggests equal impairment of performance across the two fluency tasks that is thought to reflect a generalized reduction in processing speed (Henry & Crawford, 2005).

If semantic and lexical fluency discrepancy (FD) scores are abnormal in some patients with aMCI, their magnitude and direction may prove helpful in diagnosis or prognosis. As an individually calibrated marker of performance, the direction of the discrepancy would have the advantage of being free from the need for age, gender, education, or IQ-dependent cut-off values, which require a sizable normative comparison group. Furthermore, if depressive symptoms were associated with equivalent reductions in lexical and semantic task performance (as the processing speed account would predict), then FD scores might also be of value in distinguishing between depressive and early Alzheimer related cognitive impairment. To our knowledge, no study has examined semantic and phonemic fluency in aMCI compared with healthy controls, early AD patients and patients with depressive symptoms.

Study aims

Here, we aimed to (1) replicate the finding that patients with aMCI demonstrate a pattern of fluency performance similar to that observed in patients with early AD and (2) to investigate whether their FD performance is abnormal when compared to healthy controls and out-patients with depressive symptoms.

Material and methods

Patient groups

We examined 47 patients with aMCI, 18 out-patients with depressive symptoms, 24 healthy control patients, and 35 patients with mild AD. The 47 aMCI patients were recruited over a 3-year period (September 2003–September 2006) from tertiary referrals to our neuropsychological assessment service and met criteria for aMCI (Petersen *et al.*, 1999). All aMCI patients underwent comprehensive neuropsychological and psychiatric evaluation, medical screening (including blood screen) as well as neuroimaging (CT and/or MRI or SPECT). Thirty-five patients with a NINCDS/ADRDA diagnosis (McKhann *et al.*, 1984) of early AD were identified as part of a clinical audit of service referral numbers and diagnoses. All early AD patients scored 17/30 or above on the mini mental state examination (MMSE; Folstein, Folstein, & McHugh, 1975) and 58/100 or above on the more comprehensive Addenbrooke's cognitive examination (ACE; Mathuranath, Nestor, Berrios, Rakowicz, & Hodges, 2000) indicating a relatively mild disease severity. As healthy controls we recruited 24 spouses or carers of patients who had attended the neuropsychological assessment service. Participants with potentially

confounding co-morbid medical, psychiatric, or neurological conditions (i.e. stroke or cerebro-vascular disease, head injury, alcoholism, schizophrenia, etc.) were excluded.

In an attempt to ensure a control group of similar illness severity and general level of functioning to the aMCI patients, we recruited 18 out-patients with depressive symptoms from hospital out-patient clinics and day hospitals, who were receiving the same level of out-patient care as our aMCI group. All participants in this group presented with depressive symptoms, thought not to be primarily organic in nature, yet known to have effects on cognitive functioning both during illness and after recovery (Herrmann, Goodwin, & Ebmeier, 2007). We considered matching of illness severity to be important, as in clinical practice the differentiation of severe depression and early dementia states is less problematic than separating the sequelae of the milder forms of these disorders. Furthermore, fluency measures have been shown to be sensitive to even mild depressive symptoms (Ravdin, Katzen, Agrawal, & Relkin, 2003). We included patients with a variety of disorders, as the type of depression does not appear to influence the magnitude of cognitive deficit (Christensen *et al.*, 1997). Patients with any co-morbid medical, neurological, or psychiatric condition with the potential to affect cognitive function were excluded. The mean geriatric depression scale score (Yesavage, 1988) for this group was 14.3 ($SD = 7.79$) indicating mild, yet clinically significant levels of depressive symptoms.

Neuropsychological measures

Verbal fluency

All patients were given two versions of the verbal fluency task. In the lexical version, patients were asked to generate as many words as possible within 1 minute beginning with the letter 'P'. The letter 'P' was chosen in place of the more widely known 'F', 'A', 'S' as it forms part of the ACE (Mathuranath *et al.*, 2000). In the second task, patients were asked to provide as many animal names as possible in 1 minute. Patients' scores were *z*-transformed using control means and standard deviations to be able to compare lexical and semantic task performance.

Episodic memory

All groups were administered the Hopkins verbal learning test – revised (HVLT-R; Brandt, 1991). Participants are required to recall as many words as possible immediately following presentation of a 12-item word list across three consecutive learning trials. Measures include: total number of words recalled across three registration trials (max = 36); total number of words recalled following a 30-minute delay (max = 12); and a discrimination index score representing a participant's ability to discriminate between old and new list items. The total number of words recalled immediately following list presentation provides a measure of new learning ability, while the number of words recalled after a short delay serves as a measure of delayed verbal recall ability.

The paired-associate learning (PAL) subtest from the Cambridge automated neuropsychological test assessment battery (CANTAB; Swainson *et al.*, 2001) is a computerized measure of visuospatial learning requiring participants to learn the locations of an increasing number (i.e. 1, 2, 3, 6, and then 8) of abstract patterns (Blackwell *et al.*, 2004; Swainson *et al.*, 2001). The score of interest was the number of pattern-position errors made at the six-pattern level. Healthy and depressed

control participants, as well as aMCI and a subgroup of early AD patients also completed the Rey complex figure (RCF; copy, immediate and delayed recall trials; Rey, 1941). In this task, participants are given a complex figure and are asked to make a copy of it without time restriction. Immediately after presenting the figure, and again following a 30-minute delay, participants are required to reproduce the figure from memory. The measure of interest for this analysis was the 30-minute delayed recall condition.

Attention and executive function

All participant groups completed the trail making tests, Parts A and B (TMT A and B; Reitan, 1985). Participants were required to join numbered circles in ascending order (Part A) and numbers and letters in ascending alternating sequence (Part B) at pace and the time to completion was recorded. TMT A serves as a measure of psychomotor processing speed, while TMT B also adds a divided attention component. By subtracting the time to completion for TMT A of this test from TMT B, a measure of the 'executive' functioning component can be acquired independently of processing speed.

Statistics

Data were analysed using SPSS 14.0 for Windows. A *FD score* was calculated for each participant by subtracting lexical fluency (*P words*) from semantic fluency (*animal words*). Demographic variables and cognitive performance were examined using one-way groupwise ANOVAs. As the group sizes were not equal, Tukey pairwise comparisons were carried out on all significant analyses when the assumption of homogeneity of variance was met, and Games–Howell pairwise comparisons were carried out when this assumption was violated. Reaction time data (e.g. TMT A and B) was log transformed prior to analysis in order to increase normality.

Results

Participant characteristics

Demographic characteristics are presented in Table 1. Groups did not differ in terms of age [$F(3, 120) = 1.7$; $p = .17$] or predicted pre-morbid IQ [$F(3, 105) = 2.2$; $p = .10$]. The estimated pre-morbid level of general intellectual functioning fell within a high average range for all four groups. There was a significant gender imbalance (F > M) in the depressive symptom group only ($\chi^2 = 5.6$; $p = .02$; in all other cases $p > .05$). Patients with early stage AD performed at a lower level than all other groups on cognitive screening measures (MMSE: [$F(3, 120) = 30.6$; $p < .0001$]; ACE: [$F(3, 120) = 30.6$; $p < .0001$]); in all cases $p < .0001$. Total ACE scores of the aMCI group fell between that of the controls and early AD (MCI vs. C: $p < .0001$; MCI vs. AD: $p < .0001$) and was significantly different from both of these groups. By contrast, mean MMSE scores for control and aMCI groups did not differ. Mean total scores on both cognitive screening measures (MMSE and ACE) failed to discriminate the depressive control group from aMCI or healthy control participants.

In keeping with proposed aMCI criteria (Petersen *et al.*, 1999) patients with aMCI performed more than one standard deviation below age means on at least two formal

Table 1. Means (*SD*) of clinical and demographic data

Variable	Controls (C; $N = 24$)	Depressed (D; $N = 18$)	aMCI (MCI; $N = 47$)	Early AD (AD; $N = 35$)	Statistic	*Post hoc* group differences (Tukey; Games–Howell) $p < .05$
Age	70.8 (7.8)	73.3 (6.3)	73.9 (6.4)	74.7 (6.6)	$F(3, 120) = 1.7$; $p = .17$	
Gender	9 M:15 F	4 M:14 F*	18 M:29 F	17 M:18 F	$\chi^2 = 3.5$; $p = .32$	
NART	118.5 (3.3)	116.2 (5.6)	116.8 (7.7)	113.5 (8.9)	$F(3, 105) = 2.1$; $p = .10$	
MMSE	28.9 (1.1)	28.5 (1.5)	28.3 (1.5)	24.9 (2.8)	$F(3, 120) = 30.6$; $p < .0001$	C, D, MCI > AD
ACE	94.5 (3.2)	91.3 (5.7)	89.4 (5.6)	75.5 (7.4)	$F(3, 120) = 64.5$; $p < .0001$	C > MCI C, D, MCI > AD

AD, Alzheimer's disease; aMCI, amnesic mild cognitive impairment; M, male; F, female; NART, national adult reading test; MMSE, mini mental state exam; ACE, Addenbrooke's cognitive examination.

measures of episodic verbal and visual memory, such as the HVLT total and delayed recall, the PAL task from the CANTAB, and the RCF delayed recall.

Fluency performances according to patient group

Semantic fluency

Mean scores and standard deviations for semantic fluency, lexical fluency, and semantic-lexical discrepancy are presented for each of the four patient groups in Table 2. As predicted, the mean semantic fluency score for the early AD group was significantly lower than all other patient groups (in all cases $p < .0001$). The aMCI patient group also generated significantly fewer animal names than did healthy controls ($p < .0001$). The mean number of animals generated by both the aMCI and depressed groups fell between that of early AD and control participants and was significantly different from both of these groups (MCI vs. C: $p < .0001$; MCI vs. AD: $p < .0001$; D vs. C: $p = .04$; D vs. AD: $p < .0001$).

Lexical fluency

Lexical fluency scores were higher for the aMCI and healthy control groups than for the early AD patients (C vs. AD: $p = .02$; MCI vs. AD: $p = .002$). No other group differences in lexical fluency performance were observed.

Fluency discrepancy scores

The mean FD scores for aMCI and early AD patients were significantly higher than those of healthy control (MCI vs. C: $p < .0001$; AD vs. C: $p < .0001$) and depressed control groups (MCI vs. D: $p = .01$; AD vs. D: $p = .001$), but did not differ significantly from each other (MCI vs. AD: $p = .72$) (see Figure 1).

The negative mean discrepancy scores of the aMCI and early AD groups indicate that, on average, these patients generated fewer animals than 'P' words within the 1-minute time frame. The opposite pattern (animals > P words) was observed for the healthy controls and the depressive control group, indicated by a positive mean score. Mean FD scores did not differ between the two control groups (C vs. D: $p = .27$).

To assess the usefulness of FD scores in classifying patients in comparison to a more commonly used word list learning task, receiver operating characteristic (ROC) curves were constructed (Figure 2a–c). Figure 2a and b demonstrate that the FD index has greater sensitivity and specificity in distinguishing healthy and depressed control participants from those with aMCI, whereas delayed verbal recall ability retains superiority in the early AD versus healthy control comparison (Figure 2c). The area under the curve for each of the three comparisons is as follows: C versus MCI: 0.84 (FD) and 0.78 (HVLT-delay); D versus MCI: 0.75 (FD) and 0.71 (HVLT-delay); C versus AD: 0.85 (FD) and 0.94 (HVLT-delay).

Episodic memory

By definition, the aMCI group performed at significantly lower levels than healthy controls on a number of verbal and visual episodic memory measures: HVLT total and delayed recall scores (total score [means (*SD*) for controls/patients]: 23.4 (5.1)/19.3 (4.6); $t(2, 67) = 3.4$, $p = .01$ and delayed recall: 8.1 (2.7)/4.8 (3.3); $t(2, 67) = 4.2$, $p < .001$);

Table 2. Means (and *SD*) for verbal fluency and other relevant measures of episodic lexical memory, processing speed, and executive function

Variable		Controls (C; $N = 24$)	Depressed (D; $N = 18$)	aMCI (MCI; $N = 47$)	Early AD (AD; $N = 35$)*	Statistic	*Post hoc* group differences (Tukey, Games–Howell) $p < .05$
Animal fluency	Raw score	21.3 (5.8)	17.6 (4.8)	15.4 (4.3)	10.4 (3.3)	$F(3, 120) = 30.1$; $p < .0001$	C > MCI, D > AD
	z score	0.00 (1.00)	−0.65 (0.84)	−1.03 (0.75)	−1.89 (0.57)		
P word fluency	Raw score	15.7 (5.7)	14.3 (5.4)	15.9 (4.3)	11.9 (4.4)	$F(3, 120) = 5.2$; $p = .002$	C, MCI > AD
	z score	0.00 (1.00)	−0.23 (0.95)	0.04 (0.75)	−0.66 (0.76)		
Fluency discrepancy	*z score*	0.00 (0.86)	−0.42 (0.69)	−1.07 (0.67)	−1.24 (0.79)	$F(3, 120) = 17.1$; $p < .0001$	C, D > MCI, AD
HVLT-R delay	Raw score	8.1 (2.7)	7.4 (3.6)	4.8 (3.3)	1.4 (2.8)	$F(3, 103) = 19.8$; $p < .0001$	C, D > MCI > AD
RCF delay	Raw score	16.9 (6.8)	15.3 (6.4)	11.4 (6.9)	2.4 (3.9)	$F(3, 96) = 16.3$; $p < .0001$	C, D, MCI > AD C > MCI
PAL errors	Raw score	7.9 (6.7)	11.3 (7.6)	17.1 (14.5)	40.8 (11.1)	$F(3, 109) = 38.9$; $p < .0001$	AD > C, D, MCI MCI > C
TMT A^{+}(s)	Raw score	40.2 (10.5)	53.3 (22.6)	45.1 (18.9)	59.0 (22.2)	$F(3, 104) = 4.2$; $p = .007$	AD > C, MCI
TMT B^{+}(s)	Raw score	88.7 (30.7)	140.2 (53.7)	117.1 (66.1)	183.3 (96.4)	$F(3, 106) = 8.9$; $p < .0001$	AD > C, MCI D > C
TMT B–TMT A^{+} (s)	Raw score	48.5 (24.6)	86.9 (49.1)	72.0 (60.4)	92.0 (56.8)	$F(3, 99) = 2.8$; $p = .046$	

AD, Alzheimer's disease; aMCI, amnesic mild cognitive impairment; HVLT-R delay, Hopkins verbal learning test-revised delayed recall; RCF delay, Rey complex figure delayed recall; PAL error, total number of errors made at the six pattern stage of the paired associates learning test from the CANTAB; TMT, trail making test.
*AD group sample size varies by analysis; $^{+}$Log transformed in order to increase normality of data set.

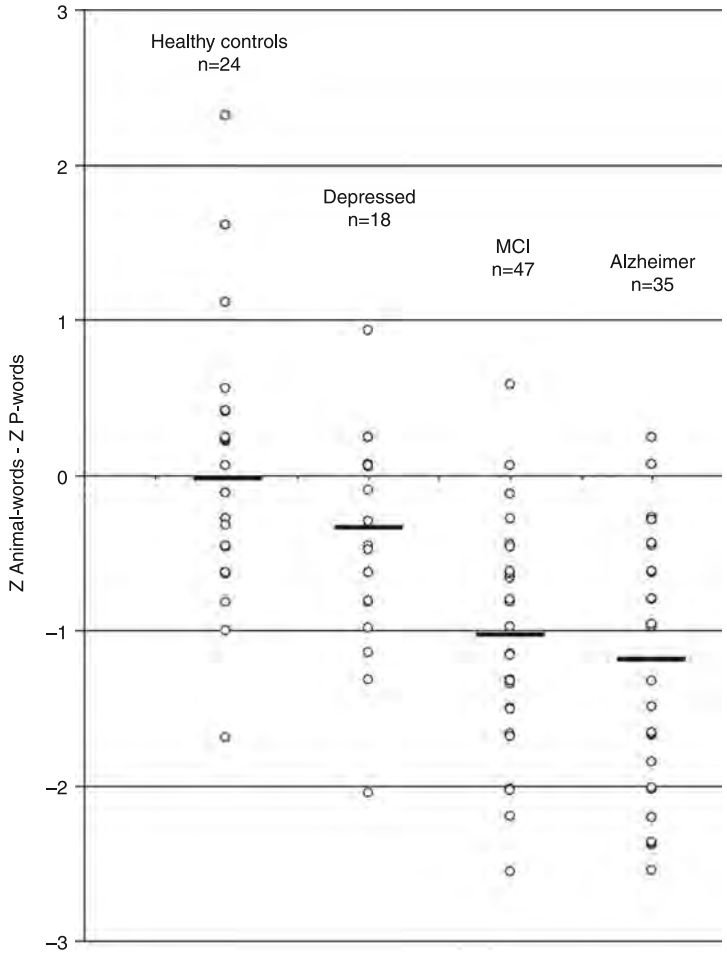


Figure 1. Means and distribution of the differences (animal minus P-word) of verbal fluency *z scores* in diagnostic groups.

the PAL task from the CANTAB (7.9 (6.7)/17.1 (14.5); $t(2, 68) = -3.7, p < .001$); and the RCF delayed recall (16.8 (6.8)/11.4 (6.9); $t(2, 66) = 3.1, p = .003$).

Attention and executive function

The early AD group were slower than healthy controls and aMCI patients to complete the TMT A (C vs. AD: $p = .01$; MCI vs. AD: $p = .03$). Furthermore, early AD and depressive control patients were on average slower than healthy controls to complete Part B of the TMT (C vs. AD: $p < .0001$; C vs. D: $p = .02$). A TMT B–A measure was calculated by subtracting the time to completion for Part A of the TMT from Part B of the same task. On analysis, there proved to be a marginally significant group difference ($F(3, 99) = 2.8$; $p = .046$); however none of the multiple comparison procedures reached significance.

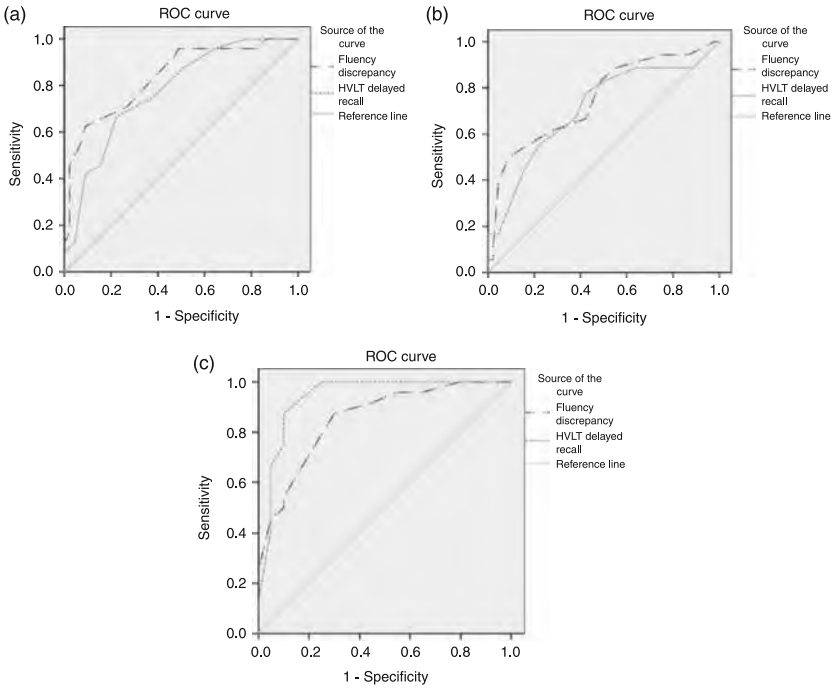


Figure 2. (a) ROC curve for differentiating aMCI participants from healthy controls. FD: AUC = 0.84, $p < .0001$; HVLT delayed recall: AUC = 0.78, $p < .0001$. (b) ROC curve for differentiating aMCI participants from depressed controls. FD: AUC = 0.75, $p = .002$; HVLT delayed recall: AUC = 0.71, $p = .009$. (c) ROC curve for differentiating Alzheimer's disease participants from healthy controls. FD: AUC = 0.86, $p < .0001$; HVLT delayed recall: AUC = 0.94, $p < .0001$.

Discussion

Our findings demonstrate that the pattern of performance on lexical and semantic fluency tasks is distinctly different in early AD and aMCI compared with healthy and depressive age-matched controls. Specifically, early AD and aMCI patients show a greater magnitude of impairment in semantic, as compared to lexical fluency, relative to healthy age-matched controls (Lezak, Howieson, Loring, Hannay, & Fischer, 2004; Spreen & Strauss, 1998).

The findings are consistent with longitudinal data demonstrating that semantic fluency deficits pre-date the lexical fluency impairment seen in the preclinical stages of AD (Auriacombe *et al.*, 2006) and with cross-sectional data showing an impairment of semantic fluency in the absence of any lexical fluency deficit in aMCI (Alladi *et al.*, 2006; Dudas *et al.*, 2005; Lonie *et al.*, 2008; Murphy *et al.*, 2006). In cases where a lexical fluency deficit has been documented in aMCI or preclinical AD, the general level of cognitive function of the aMCI patients is often comparatively low for example mean MMSE of 24.5, compared to our early AD patient group of 24.9 (Jones *et al.*, 2006).

Our results implicate early failure of one or more aspect(s) of the semantic memory system in aMCI and early AD. They do not directly address the question of which

mechanisms of the semantic memory system may be at fault. The relatively sound performance of our aMCI patients on additional measures of verbal initiation and speeded divided attention would suggest that these specific aspects of executive function do not underlie the semantic fluency deficit. Similarly, Auriacombe and colleagues (Auriacombe *et al.*, 2006) failed to implicate executive processes in the verbal fluency decline characterizing AD. It remains possible that the semantic fluency deficit seen in aMCI and early AD reflects degradation of amodal representations forming the semantic memory store, or failure of the executive processes that help to direct and control semantic activation of this store (Jefferies & Lambon Ralph, 2006). Detailed quantitative and qualitative longitudinal investigation of semantic memory function in an aMCI cohort will be required to differentiate between these possibilities. What appears increasingly clear, nonetheless, is the coexistence of cognitive deficits in domains other than episodic memory function in aMCI; most notably semantic memory deficits (Adlam, Bozeat, Arnold, Watson, & Hodges, 2006; Alladi *et al.*, 2006; Lam, Ho, Lui, & Tam, 2006; Murphy *et al.*, 2006; Perry & Hodges, 2000; Ribeiro, de Mendonca, & Guerreiro, 2006; Vogel, Gade, Stokholm, & Waldemar, 2005). The original MCI criteria have been modified to some extent to reflect this (Petersen, 2004).

FD scores performed as well as, if not better than delayed verbal recall measures in distinguishing between patients with aMCI or early AD- and age-matched controls or depressive controls (ROC). The discrepancy score may therefore be especially sensitive to semantic memory failure in aMCI. There are a number of reasons why this might be so.

Verbal fluency scores are influenced by age and IQ (Lezak *et al.*, 2004; Spreen & Strauss, 1998; Tombaugh, Kozak, & Rees, 1999), and performance on the two types of fluency measures is known to correlate (Tombaugh *et al.*, 1999). Applying generalized cut-off scores at an individual patient level could conceivably reduce the sensitivity of fluency measures to deficits of a small magnitude. Lexical fluency performance, which remains comparatively well preserved in preclinical and very early AD (Alladi *et al.*, 2006; Dudas *et al.*, 2005; Murphy *et al.*, 2006) and is relatively insensitive to the effects of ageing in later life (Murphy *et al.*, 2006), may act as a personalized benchmark, against which even subtle declines in an individual's semantic fluency performance can be measured. At least one study has shown an enhanced ability to detect cognitive deficits of a progressive nature using individualized IQ adjusted rather than group norms (Rentz *et al.*, 2004). Based on a similar rationale is the use of regression equations to calculate expected performance levels for the individual patient (Crawford & Garthwaite, 2006; Rentz *et al.*, 2004).

An interesting observation was the comparative lesser ability of the FD score to discriminate early AD sufferers from healthy age-matched controls. Prior research with established AD sufferers suggests that the relative magnitude of phonemic and semantic fluency deficits remains constant across disease stages (Henry *et al.*, 2004). This would imply that the discriminative power obtained for FD in the aMCI group should also be maintained in early AD. It is possible however that this argument does not hold in a preclinical disease phase; at which time the greater sensitivity of the discrepancy score may reflect semantic memory failure prior to disruption of wider executive processes.

There is continuing debate as to which of the two cognitive domains, that is semantic memory abilities or executive processes, constitutes the secondary area of impairment in early and preclinical AD. It is conceivable that as the disease progresses, 'executive' aspects of semantic retrieval may add to poor performance due to existing semantic memory dysfunction on category fluency tasks. When this occurs, deficits on

lexical tasks would also be expected to be present. As lexical fluency performance declines, the discrepancy in fluency performance may become a less sensitive measure of cognitive dysfunction than category fluency performance or delayed recall alone.

In our study, the differential semantic fluency deficit was unique to early AD and aMCI patient groups suggesting that it may be of assistance in differentiating these groups from the healthy elderly or elderly patients with depressive symptoms. Other studies have found that comparison of performance on the two different types of fluency tests is helpful in distinguishing between certain dementia syndromes (Jones *et al.*, 2006; Marczinski & Kertesz, 2006). This level of discrimination may not be achievable by examination of fluency scores in isolation, or indeed by cognitive screening measures, both of which failed to discriminate between our aMCI and depressive control groups.

The use of a single letter and category (i.e. P and animals), as opposed to an average of three (i.e. FAS or animals, fruits, and vegetables), might be expected to yield a less reliable fluency score, and it will be necessary to replicate the above findings in other clinical samples using alternative categories and letters. One recent study, comparing the number of animals with F-words generated, has demonstrated a similar ability to discriminate early AD and aMCI patients from healthy controls. This study lacked the important inclusion of a depressive control group matched for general level of cognitive functioning and support needs (Murphy *et al.*, 2006). Furthermore, the advantage of the measures used in the current study is that they are obtainable within the context of a brief cognitive screening measure (ACE and ACE-R), without necessity for the administration of any supplementary tests.

In order for FD scores to be of value in a differential diagnostic sense, they must facilitate discrimination between pathological and age appropriate performance at the individual patient level. Only one healthy control and one out-patient with depressive symptoms in our study obtained a higher lexical than category fluency score, but there was a good deal of directional variability in the early AD and aMCI patient groups. Hence the FD measure may prove to be of greatest assistance identifying those persons who are not likely to develop AD over time (negative predictive value). We are conducting longitudinal follow-up of the aMCI patients in order to determine whether or not this is the case.

Two previous studies have contrasted fluency difference scores in AD and healthy controls (Cerhan *et al.*, 2002; Sherman & Massman, 1999). In both, the overwhelming majority of AD patients demonstrated the expected semantic $<$ lexical fluency performance pattern. However, the existence of a notable number of AD patients exhibiting the opposite pattern led both authors to conclude that the FD score may be of limited sensitivity.

In the present study, optimal cut-off scores can be generated from the receiver–operator curves in Figure 2. For example, a cut-off of 3.5 and higher for the animal minus letter-P scores (or a difference of *z scores* of -37%) gives a sensitivity of 63 and 50% to aMCI in healthy volunteers and participants with depressive symptoms, respectively, but a very high specificity of 91% (Table 3). On the other hand, a more lenient cut-off of greater or equal to 1.5 (*z score* of -72%) gives a sensitivity for aMCI or Alzheimer's dementia versus controls of 88%, with a reduced specificity of 58 and 70%, respectively.

Our results indicate that the differential semantic fluency deficit seen in AD is also present in aMCI. They provide support for the presence of semantic memory impairment in the preclinical and very early stages of AD and emphasize the need to broaden the current conceptualization of aMCI beyond that of a purely amnesic state.

Table 3. Optimum cut-off values for animal minus letter fluency to achieve high sensitivity and specificity, respectively

	Difference of raw scores	Difference of *z scores* [%]	Sensitivity [%]	Specificity [%]
MCI versus controls	≥3.5	≥ − 37	63	91
	≥1.5	≥ − 72	88	58
MCI versus depressed	≥3.5	≥ − 37	50	91
	≥ − 0.5	≥ − 107	83	51
ATD versus controls	≥1.5	≥ − 72	88	70

FD scores appear equally adept, if not superior, to episodic memory measures at identifying aMCI. Importantly, the presence of depressive symptoms in psychiatric outpatients does not appear to influence the relative performance on the two types of fluency measures, suggesting that consideration of FD may be of assistance in early differential diagnosis. Longitudinal follow-up of the aMCI will determine whether or not FD scores are of equal utility in a prognostic sense.

Acknowledgements

This study was financially supported by the Gordon Small Charitable Trust and the European Commission Network of Excellence Diagnostic Molecular Imaging (FP6-LIFESCIHEALTH Project Reference 512146). We thank all the participants and our clinical colleagues for referring their patients.

References

Adlam, A. L. R., Bozeat, S., Arnold, R., Watson, P., & Hodges, J. R. (2006). Semantic knowledge in mild cognitive impairment and mild Alzheimer's disease. *Cortex*, *42*(5), 675–684.

Alladi, S., Arnold, R., Mitchell, J., Nestor, P. J., & Hodges, J. R. (2006). Mild cognitive impairment: Applicability of research criteria in a memory clinic and characterization of cognitive profile. *Psychological Medicine*, *36*(4), 507–515.

Auriacombe, S., Lechevallier, N., Amieva, H., Harston, S., Raoux, N., & Dartigues, J. F. (2006). A longitudinal study of quantitative and qualitative features of category verbal fluency in incident Alzheimer's disease subjects: Results from the PAQUID study. *Dementia and Geriatric Cognitive Disorders*, *21*(4), 260–266.

Beatty, W. W., Salmon, D. P., Troster, A. I., & Tivis, R. D. (2002). Do primary and supplementary measures of semantic memory predict cognitive decline by patients with Alzheimer's disease? *Aging, Neuropsychology, and Cognition*, *9*(1), 1–10.

Blackwell, A. D., Sahakian, B. J., Vesey, R., Semple, J. M., Robbins, T. W., & Hodges, J. R. (2004). Detecting dementia: Novel neuropsychological markers of preclinical Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, *17*(1–2), 42–48.

Brandt, J. (1991). The Hopkins verbal learning test: Development of a new memory test with six equivalent forms. *Clinical Neuropsychologist*, *5*(2), 125–142.

Cerhan, J. H., Ivnik, R. J., Smith, G. E., Tangalos, E. C., Petersen, R. C., & Boeve, B. F. (2002). Diagnostic utility of letter fluency, category fluency, and fluency difference scores in Alzheimer's disease. *Clinical Neuropsychologist*, *16*(1), 35–42.

Christensen, H., Griffiths, K., Mackinnon, A., & Jacomb, P. (1997). A quantitative review of cognitive deficits in depression and Alzheimer-type dementia. *Journal of the International Neuropsychological Society*, *3*(6), 631–651.

Crawford, J. R., & Garthwaite, P. H. (2006). Comparing patients' predicted test scores from a regression equation with their obtained scores: A significance test and point estimate of abnormality with accompanying confidence limits. *Neuropsychology*, *20*(3), 259–271.

Dudas, R. B., Clague, F., Thompson, S. A., Graham, K. S., & Hodges, J. R. (2005). Episodic and semantic memory in mild cognitive impairment. *Neuropsychologia*, *43*(9), 1266–1276.

Folstein, M., Folstein, S., & McHugh, P. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatry Research*, *12*, 189–198.

Grundman, M., Petersen, R. C., Ferris, S. H., Thomas, R. G., Aisen, P. S., Bennett, D. A., *et al.* (2004). Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Archives of Neurology*, *61*(1), 59–66.

Henry, J. D., & Crawford, J. R. (2005). A meta-analytic review of verbal fluency deficits in depression. *Journal of Clinical and Experimental Neuropsychology*, *27*(1), 78–101.

Henry, J. D., Crawford, J. R., & Phillips, L. H. (2004). Verbal fluency performance in dementia of the Alzheimer's type: A meta-analysis. *Neuropsychologia*, *42*(9), 1212–1222.

Herrmann, L. L., Goodwin, G. M., & Ebmeier, K. P. (2007). The cognitive neuropsychology of depression in the elderly. *Psychological Medicine*, *37*(12), 1693–1702.

Jefferies, E., & Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia versus semantic dementia: A case-series comparison. *Brain*, *129*(8), 2132–2147.

Jones, S., Laukka, E. J., & Backman, L. (2006). Differential verbal fluency deficits in the preclinical stages of Alzheimer's disease and vascular dementia. *Cortex*, *42*(3), 347–355.

Lam, L. C., Ho, P., Lui, V. W., & Tam, C. W. (2006). Reduced semantic fluency as an additional screening tool for subjects with questionable dementia. *Dementia and Geriatric Cognitive Disorders*, *22*(2), 159–164.

Lezak, M. D., Howieson, D. B., Loring, D. W., Hannay, H. J., & Fischer, J. S. (2004). *Neuropsychological assessment*. New York: Oxford University Press.

Lonie, J. A., Herrmann, L. L., Donaghey, C. L., & Ebmeier, K. P. (2008). Mild cognitive impairment referral patterns and cognitive profile: Some implications for clinical practice. *British Journal of Psychiatry*, *192*, 1–7. doi: 10.1192/bjp.bp. 107.035642.

Marczinski, C. A., & Kertesz, A. (2006). Category and letter fluency in semantic dementia, primary progressive aphasia, and Alzheimer's disease. *Brain and Language*, *97*(3), 258–265.

Mathuranath, P. S., Nestor, P. J., Berrios, G. E., Rakowicz, W., & Hodges, J. R. (2000). A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. *Neurology*, *55*(11), 1613–1620.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS–ADRDA work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology*, *34*(7), 939–944.

Murphy, K. J., Rich, J. B., & Troyer, A. K. (2006). Verbal fluency patterns in amnestic mild cognitive impairment are characteristic of Alzheimer's type dementia. *Journal of the International Neuropsychological Society*, *12*(4), 570–574.

Perry, R. J., & Hodges, J. R. (2000). Fate of patients with questionable (very mild) Alzheimer's disease: Longitudinal profiles of individual subjects' decline. *Dementia and Geriatric Cognitive Disorders*, *11*(6), 342–349.

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, *256*(3), 183–194.

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology*, *56*(3), 303–308.

Ravdin, L. D., Katzen, H. L., Agrawal, P., & Relkin, N. R. (2003). Letter and semantic fluency in older adults: Effects of mild depressive symptoms and age-stratified normative data. *Clinical Neuropsychologist*, *17*(2), 195–202.

Reitan, R. M. (1985). *Halstead–Reitan neuropsychological test battery.* Tucson: Reitan Neuropsychology Laboratory/Press.

Rentz, D. M., Huh, T. J., Faust, R. R., Budson, A. E., Scinto, L. F., Sperling, R. A., *et al.* (2004). Use of IQ-adjusted norms to predict progressive cognitive decline in highly intelligent older individuals. *Neuropsychology*, *18*(1), 38–49.

Rey, A. (1941). Psychological examination of a case of post-traumatic encephalopathy. *Archives de Psychologie*, *28*, 286–340.

Ribeiro, F., de Mendonca, A., & Guerreiro, M. (2006). Mild cognitive impairment: Deficits in cognitive domains other than memory. *Dementia Geriatric Cognitive Disorders*, *21*(5–6), 284–290.

Sherman, A. M., & Massman, P. J. (1999). Prevalence and correlates of category versus letter fluency discrepancies in Alzheimer's disease. *Archives of Clinical Neuropsychology*, *14*(5), 411–418.

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests. Administration, norms and commentary* (2nd ed.). New York: Oxford University Press.

Swainson, R., Hodges, J. R., Galton, C. J., Semple, J., Michael, A., Dunn, B. D., *et al.* (2001). Early detection and differential diagnosis of Alzheimer's disease and depression with neuropsychological tasks. *Dementia Geriatric Cognitive Disorders*, *12*(4), 265–280.

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology*, *14*(2), 167–177.

Vogel, A., Gade, A., Stokholm, J., & Waldemar, G. (2005). Semantic memory impairment in the earliest phases of Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, *19*(2–3), 75–81.

Yesavage, J. A. (1988). Geriatric depression scale. *Psychopharmacological Bulletin*, *24*(4), 709–711.

Zakzanis, K. K., Leach, L., & Kaplan, E. (1998). On the nature and pattern of neurocognitive function in major depressive disorder. *Neuropsychiatry, Neuropsychology and Behavioural Neurology*, *11*(3), 111–119.

Received 19 September 2007; Received in revised form 4 December 2007

BJPsych | The British Journal of Psychiatry (2008) 192, 59–64. doi: 10.1192/bjp.bp.107.035642

Clinical referral patterns and cognitive profile in mild cognitive impairment

Jane A. Lonie, Lucie L. Herrmann, Claire L. Donaghey and Klaus P. Ebmeier

Background
There is current interest in exploring the different subtypes of mild cognitive impairment (MCI), in terms of both their epidemiology and their cognitive profile.

Aims
To examine the frequency of MCI subtypes presenting to a memory clinic and to document detailed neuropsychological profiles of patients with the amnestic subtype.

Method
Consecutive tertiary referrals (*n*=187) were psychiatrically evaluated; 45 patients met criteria for amnestic mild cognitive impairment (aMCI). A subgroup of 33 patients with aMCI as well as 21 healthy controls took part in a thorough neuropsychological examination.

Results
Of the patients who were examined in greater neuropsychological detail, ten had pure aMCI (none with visual memory impairment only). Fifteen met criteria for non-amnestic MCI. Fifteen had normal neuropsychological profiles. Using more than one test increased sensitivity to detect episodic memory impairment.

Conclusions
Amnestic MCI is an important diagnosis in secondary and tertiary memory clinics. There is scope to improve the efficacy and sensitivity of the clinical assessment of this impairment.

Declaration of interest
None.

The diagnosis of mild cognitive impairment (MCI) represents an attempt to define features of the dementias in their preclinical phases. Petersen and colleagues developed research criteria with a cut-off for verbal recall performance as objective evidence of episodic memory impairment.[1] This approach has been challenged because it excludes patients who display exclusively visual episodic memory impairment.[2] Moreover, the pure amnestic subtype of mild cognitive impairment is rare,[2–6] and mild cognitive impairment case definition varies as a function of the neuropsychological tests used.[2,7] It is not clear how the Petersen criteria might best be translated into clinical practice. There is little information detailing the frequency with which each of the subtypes presents to memory clinics. Our aim was to examine the diagnostic profile of patients with amnestic mild cognitive impairment (aMCI) referred to our tertiary assessment service. We also sought to evaluate a comprehensive battery of neuropsychological measures for their usefulness in this patient group.

Method

This study constituted part of a longitudinal project for which ethical approval was obtained from the local research ethics committee. In accordance with this, informed consent was given by all participants.

Sample

We retrospectively analysed 187 consecutive referrals to the Edinburgh Neuropsychological Assessment Service for Older People between the months of September 2004 and April 2006. Referrals were received at a tertiary level, stemming from consultants in old age psychiatry, geriatric medicine and neurology. All of these patients had undergone comprehensive psychiatric evaluation, relevant medical screening (including a standard battery of screening blood tests) and neuroimaging (computed tomography and/or magnetic resonance imaging or single-proton emission computed tomography) prior to being referred to our service. All but three patients were over the age of 50 years.

The original criteria for mild cognitive impairment set out by Petersen *et al* require that a person must present with a memory complaint, show evidence of objective memory decline in relation to age and education, demonstrate preservation of other areas of cognitive function and activities of daily life, and not fulfil criteria for dementia.[8] Because it has since become apparent that not everyone who demonstrates cognitive impairment short of dementia has a 'memory' complaint, we used the recently expanded criteria that include people with non-memory complaints (single-domain non-memory MCI), as well as those exhibiting multiple domains of cognitive impairment who none the less fail to fulfil criteria for dementia (multiple domains slightly impaired).[9,10] The Mini Mental State Examination (MMSE)[11] and Addenbrooke's Cognitive Examination[12] were administered as a means of establishing the participants' general level of cognitive functioning. Level of everyday functioning was examined by means of the Clinical Dementia Rating scale[13] within the context of a clinical interview with the patient and the patient's primary carer (when available). A total of 112 patients fulfilled one or more of the following exclusion criteria and were therefore excluded from the analyses:

(a) dementia (MMSE score $<24/30$ or ACE score $<80/100$, plus fulfilling DSM–IV criteria;[14]

(b) depression, assessed either by way of formal psychiatric consultation or, in a small proportion of cases, by a score greater than 10 on the Geriatric Depression Scale[15] or clinical assessment by one of the authors (J.A.L.);

(c) one or more medical or psychiatric conditions that could conceivably account for the patient's cognitive impairment (head injury, schizophrenia, evidence of stroke or tumour on neuroimaging, alcoholism, epilepsy, cranial radiotherapy).

Of the remaining 75 patients, 15 showed cognitive impairments outside the domain of episodic memory, 15 returned a 'normal' cognitive profile and 45 showed memory function

impaired for age (with or without additional areas of cognitive impairment). We present detailed neuropsychological baseline findings for 33 of these 45 patients as well as for 21 healthy individuals from the community who agreed to participate as a control group in a continuing longitudinal study examining neuropsychological markers of preclinical dementia. Control group participants were recruited through a local dementia support group or were spouses or carers of patients who had attended the neuropsychological assessment service.

Neuropsychological assessment measures

All the participants were given a comprehensive battery of neuropsychological tests. These tests were selected on the basis of their demonstrated validity for use within a population with MCI, and assessed the primary domains of verbal and visual episodic memory, semantic memory and language, processing speed, attention/executive function and visuospatial ability.

Premorbid intellectual ability

The National Adult Reading Test (NART)[16] was administered in order to assess probable premorbid level of intellectual function.

Episodic memory

To assess verbal episodic memory, participants were given the Hopkins Verbal Learning Test – Revised (HVLT–R).[17] In this test participants are asked to recall as many words as possible immediately after presentation of a 12-item word list on three consecutive learning trials. Measures included total number of words recalled across three registration trials (maximum 36), total number of words recalled following a 30 min delay (maximum 12) and a discrimination index score representing a participant's ability to discriminate between old and new list items. Visual episodic memory was assessed by means of two different tasks. The CANTAB Paired Associate Learning (PAL) test is a computerised measure of visuospatial learning ability requiring participants to learn the locations of an increasing number of patterns – one, two, three, six and then eight;[18,19] the score of interest was the number of pattern-position errors at the six-pattern level. Participants were also administered the Rey Complex Figure Test.[20] For this test, participants are asked to make a copy of a complex figure, with no time restriction. Immediately after presenting the figure, and again following a 30 min delay, participants are required to make another copy from memory.

Semantic memory

Participants completed the Graded Naming Test,[21] the Graded Faces Test,[22] the Boston Naming Test[23] and the Edinburgh Exemplar Naming Test (EENT; further details available from the authors). The Graded Naming Test and Boston Naming Test require participants to name line drawings of increasing difficulty, whereas the Graded Faces Test requires participants to name a series of 30 famous faces.[22] The EENT was developed by one of the authors (J.A.L.) in an effort to improve the sensitivity of existing confrontation naming measures to early semantic memory failure. In this test the participant is required to name 50 line drawings of low-frequency animate objects with sizeable feature overlap. Participants were also asked to complete a category fluency task, requiring them to name as many animals as they could in 1 min.

Attention/executive functioning

As a means of examining attention/executive function and visuomotor processing speed, participants were asked to produce as many words as possible beginning with the letter P in 1 min (letter fluency task). In addition, participants were administered parts A and B of the Trail Making Test:[24] in this test participants are required to join up as quickly as possible numbered circles in ascending order (part A) and numbers and letters in ascending alternating sequence (part B), while the time to completion is recorded.

Visuospatial skills

Visuospatial skills were assessed by means of the Rey Complex Figure Test copy task described above.[20]

Comparison with other memory clinics

We searched the literature for studies employing neuropsychological test batteries similar to ours to examine the comparability of our sample with other published data.

Statistical analysis

We calculated *z*-scores to determine where scores fell below the tenth percentile of control performances. Visual inspection together with a one-sample Kolmogorov–Smirnov goodness-of-fit test indicated that the data were normally distributed. Group means were compared using independent sample *t*-tests. To determine whether there is an association between general level of cognitive function and the consistency of episodic memory impairment, we divided the participants with cognitive impairment into two groups: those who displayed episodic memory impairment on a single measure only and those who showed impairment on two or more episodic memory tests. Their Addenbrooke's Cognitive Examination and MMSE scores were compared using an independent samples *t*-test. Diagnostic categories between memory clinics were compared using chi-squared tests for diagnostic categories.

Results

Literature search

We identified one other study reporting consecutive referrals to a memory clinic using similar diagnostic criteria and assessment measures.[2]

Comparability of referral patterns and cognitive profiles

A striking similarity in referral patterns was observed between our Neuropsychological Assessment Service for Older Adults and data reported recently from the Cambridge Memory Clinic.[2] When the 150 pre-excluded referrals from the Cambridge Memory Clinic were accounted for, over half (60%) of referrals from both centres were excluded on the grounds of an established dementia or depressive disorder, or one or more medical conditions that could account for the patient's cognitive impairment (χ^2=0.015, P=0.90). Close to 40% of referrals from both centres fell within the non-demented and non-depressed category ($\chi^2 < 0.001$, P=0.995). Just over half of these patients in both centres met Petersen's expanded criteria[9] for aMCI (χ^2=0.46, P=0.50), representing close to a fifth of overall referrals from both centres. Of the remaining 40% of patients in the non-demented, non-depressed category, half demonstrated cognitive deficits of a non-amnestic variety (in one or more domains) and half returned 'normal' cognitive profiles. Although referral patterns for aMCI were similar

across the two centres, there was a greater proportion of patients with non-amnestic MCI and fewer with visual-only impairment in our sample (χ^2=13.23, d.f.=3, P=0.004; Fisher–Freeman–Halton exact test, P=0.003). Mean scores for both the aMCI and control groups across all neuropsychological measures were similar to those previously reported.

Sample characteristics

Our final sample for analysis consisted of 33 patients with aMCI (13 men and 20 women, with a mean age of 74.0 years, s.d.=6.4; social class I n=9, II n=12, IIIN n=9, social class not known n=3) and 21 healthy community-dwelling older adults without cognitive complaints (7 men and 14 women, with a mean age of 69.5 years, s.d.=7.4). These groups did not differ in terms of estimated premorbid level of intellectual function. The mean age of our control group was, however, significantly lower than that of our aMCI patient group, a finding similar to previous reports.[2,4]

Comparison of aMCI and control groups

Despite a mean ACE score that exceeded suggested cut-off points for dementia,[12] the aMCI group had significantly lower mean scores than the control group on all neuropsychological measures, with the exception of the Rey Complex Figure Test copy task, the Trail Making Test part A and the letter (phonemic) fluency task. These findings were confirmed when the analysis was re-run with age as a covariate, with the exception of performance on the CANTAB PAL and the Trail Making Test part B, which just failed to reach significance. Demographic data and mean scores for the two groups on the individual neuropsychological measures are provided in Table 1. All aMCI group means are based on data from 33 participants, with the exception of the final three semantic memory measures, for which the participant numbers ranged between 13 and 31.

Table 1 Demographic data and performance of the sample on our neuropsychological test battery

	Control group (n=21) Mean (s.d.)	aMCI group (n=33) Mean (s.d.)	Control *v.* aMCI[a] P
Age, years	69.8 (7.4)	74.0 (6.4)	0.02*
NART score	118.3 (2.8)	116.4 (8.5)	0.331
MMSE score (30)[b]	29.1 (0.8)	28.0 (1.8)	<0.01**
ACE score			
Total (100)	94.8 (3.3)	88.2 (5.9)	<0.001***
Delay (7)	6.4 (0.9)	3.9 (2.3)	<0.001***
Rey Complex Figure Test scores			
Copy (36)	34.0 (2.4)	34.2 (2.5)	0.729
Immediate recall (36)	19.1 (6.3)	12.3 (6.0)	<0.001**
Delayed recall (36)	17.4 (7.4)	10.8 (6.9)	<0.01**
HVLT–R score			
Total recall (12)	24.3 (5.0)	18.2 (4.6)	<0.001***
Delayed recall (12)	8.19 (2.8)	4.8 (3.2)	<0.001**
Discrimination (12)	10.3 (1.9)	8.5 (2.2)	<0.01**
PAL 6 errors	7.8 (6.9)	16.6 (14.8)	<0.01**
Trail Making Test score			
Part A, s	40.3 (11.2)	48.4 (20.0)	0.095
Part B, s	87.5 (31.6)	131.7 (78.4)	<0.01**
Animal fluency score	21.1 (5.7)	15.1 (4.5)	<0.001***
P words score	15.7 (5.8)	16.0 (4.8)	0.860
Boston Naming Test score (60)	57.4 (3.1)	53.6 (5.5)	<0.01**
EENT score (50)	46.8 (3.0)	43.6 (4.5)	<0.01**
Graded Faces Naming Test score (30)	20.8 (3.1)	16.3 (4.9)	<0.001***
Graded Naming Test score (30)	23.4 (3.2)	19.5 (4.2)	0.015*

ACE, Addenbrooke's Cognitive Examination; aMCI, amnestic mild cognitive impairment; EENT, Edinburgh Exemplar Naming Test; HVLT–R, Hopkins Verbal Learning Test – Revised; MMSE, Mini Mental State Examination; NART, National Adult Reading Test; PAL, CANTAB Paired Associates Learning test.
a. Independent sample *t*-tests comparing the two groups.
b. The maximum test score is given in parentheses after each test name in the left-hand column.
*P<0.05; P<0.01; P<0.001.

Amnestic MCI group performance

Episodic memory

The neuropsychological performances of 33 of the 45 patients who fulfilled Petersen's expanded criteria[9] for aMCI were examined in greater detail. As in the Cambridge Memory Clinic study,[2] not all of these participants demonstrated impairment across all episodic memory measures: 11 (33%) showed impairment on a single test, 9 (27%) showed impairment on two memory measures and the remaining 13 (39%) were impaired on three or more tests. Mean MMSE and Addenbrooke's Cognitive Examination scores for participants who were impaired on more than one episodic memory measure were significantly lower than for those showing impairment on a single test (P<0.05). Just over half of our aMCI subgroup showed both verbal and visual episodic memory impairment. Although a significant proportion (45%) demonstrated memory impairment of a verbal nature only, none of our patients in this group exhibited a pure visual memory deficit.

Non-memory measures

Only ten (30%) patients in our aMCI subgroup exhibited an isolated impairment of episodic memory function. All the other patients (70%) exhibited deficits in one or more additional domains of cognition, most commonly that of semantic memory function, followed by attention and executive function (Table 2).

Discussion

Referral patterns

In this study we have shown that people who are neither depressed nor demented but who fulfil Petersen's expanded criteria[9] for aMCI make up a significant proportion of referrals to our Old Age Clinical Neuropsychology service. Roughly a quarter of all patients referred during an 18-month period met Petersen's criteria, of whom a minority exhibited a memory deficit in isolation following comprehensive neuropsychological examination. This is an almost identical proportion of patients to that reported in the study by the Cambridge Memory Clinic (23%),[2] and is a very similar figure to the 20% of patients seeking help in Lehrner *et al*'s clinic who met criteria for aMCI.[25] Furthermore, the observation that the highest numbers of people with MCI were those with more than one domain of cognitive deficit is in keeping with the findings from both population-based and other memory clinic studies.[7,26]

It is possible that referral patterns might differ depending on whether the memory clinics are geriatrician- or psychiatrist-led. Surprisingly, the level at which referrals are received (whether primary, as in the Cambridge Memory Clinic, or tertiary, as in our clinic) appears not to influence the proportion of referrals of an MCI nature that are received.

Thus, it appears that both the concept and criteria are applicable and, indeed, a necessary adjunct to clinical practice.

Table 2 Performance of patients with amnestic mild cognitive impairment (n=33) on individual episodic memory measures and non-amnestic measures

Measure	Patients performing below 10th percentile of control group performance *n* (%)
Episodic memory	
HVLT–R total recall	17 (52)
HVLT–R delay	16 (48)
HVLT–R discrimination index	13 (39)
Any HVLT–R measure	24 (73)
PAL errors stage 6	15 (46)
Rey test delay	11 (33)
ACE delay	23 (70)
Patients with impairment on 1 measure only	11 (33)
Patients with impairments on 2 measures	9 (27)
Patients with impairments on 3 measures	6 (18)
Patients with impairments on 4 measures	7 (21)
Semantic memory/language	
BNT	13 (39)
EENT	11 (33)
GFT	12 (36)
GNT	4 (12)
Animal fluency	17 (52)
Participants showing impairment on one or more semantic memory measure	22 (67)
Attentional/executive function	
P words	1 (3)
TMT part B	11 (33)
Total showing impairment on one or more attentional/executive measure	11 (33)
Visuospatial function	
Rey copy	2 (6)
Visuomotor processing speed	
TMT part A	2 (6)

ACE, Addenbrooke's Cognitive Examination; aMCI, amnestic mild cognitive impairment; BNT, Boston Naming Test; EENT, Edinburgh Exemplar Naming Test; GFT, Graded Faces Test; GNT, Graded Naming Test; HVLT–R, Hopkins Verbal Learning Test – Revised; PAL, CANTAB Paired Associates Learning test; Rey, Rey Complex Figure Test; TMT, Trail Making Test.

Neuropsychological profile of aMCI

The applicability of Petersen's research criteria[1] to clinical practice has recently been challenged on the grounds of exclusion of a significant number of patients who display episodic memory impairment of a visual nature only.[2] In contrast to the results from the Cambridge Memory Clinic reported by Alladi *et al*,[2] around half of the participants with aMCI in our study showed impairment of both verbal and visual memory, whereas all of the remaining participants with aMCI exhibited memory impairment of a verbal nature only. That is to say, we failed to uncover any case of isolated visual memory impairment.

A recent study[7] supports the notion that the type of episodic memory measure used may affect whether or not impairments are detected – a point we will return to later in the discussion. However, the absence of participants showing episodic memory impairment of a solely visual nature in our sample cannot be readily explained in terms of differential test sensitivities, because near-identical neuropsychological measures were used in previous studies (e.g. that of Alladi *et al*[2]) to assess visual memory function. Only a small proportion of our patients with aMCI demonstrated impairment on the visual episodic memory tasks *per se*. Administrative procedures might go some way to explain this observation. Specifically, our inclusion of an immediate Rey Complex Figure Test recall trial might have resulted in higher delay scores,[27] thus serving to reduce the sensitivity of this measure in our aMCI group.

Findings of studies examining patients who are at risk of developing Alzheimer's disease suggest that measures of verbal episodic memory are most sensitive to changes early in the disease course, followed by measures of visual memory.[28] It is therefore conceivable, taking into account our aMCI group's higher mean score on the Addenbrooke's Cognitive Examination, that our sample contained a greater number of patients who were at an earlier stage of their disease course. Longitudinal follow-up, in particular observation of annual performances on these visual episodic memory measures, will determine whether this is indeed the case.

Several studies have drawn attention to the substantial variability in MCI case definition as a function of the specific neuropsychological tests used.[2,7] Consistent with this, in our study there was variability among the aMCI group as to which and how many episodic memory measures were impaired. This finding was previously demonstrated,[2] and highlights the inherent difficulty in specifying the use of any single measure as a means of establishing impaired episodic memory function in aMCI. Our aMCI sample could be roughly divided into thirds in terms of numbers of participants exhibiting impairment on one, two and three or more episodic memory measures. A similar breakdown in numbers has been previously reported.[2] It would appear entirely reasonable and indeed a matter of good clinical practice to seek to establish consistency in performance across a range of episodic memory measures in defining aMCI and it will be of interest to see whether this is a significant determinant of outcome.

The variability in case definition of aMCI as a function of the cognitive measures employed, coupled with the inherent difficulties in specifying the use of a single common measure in the evaluation of this condition, poses a major challenge for clinicians. Our findings suggest that employing Petersen's expanded criteria[9] for MCI could conceivably lead to a patient's condition being classified as single-domain aMCI, multiple-domain aMCI or 'worried well', depending on the cognitive measures that were employed. If the MCI subdivisions prove useful in a prognostic sense, the means by which the cognitive aspects of the criteria are put into operation by clinicians will require further clarification.

Mean scores on cognitive screening measures were significantly lower for participants showing impairment on more than one episodic memory measure. This may reflect a more advanced disease course in this group. It is also possible that the single-measure impairment group will prove to be a less stable one over time, with a number of patients returning normal neuropsychological profiles when tested again at a later date. Alternatively, in cases in which participants show impairment on a single verbal memory measure only, this might have arisen secondarily to impairment in another cognitive domain, for example expressive language or attention/executive function (in which case the person's condition might be more accurately conceptualised as non-amnestic MCI). These possibilities and the prognostic implications of consistency and pervasiveness of impaired episodic memory performances remain to be examined by way of longitudinal follow-up.

Our study adds to the growing body of evidence supporting the rarity of a pure amnestic MCI syndrome,[2,5,6] and demonstrates that additional impairment often goes unnoticed unless participants undergo thorough neuropsychological assessment. Among 33 patients with aMCI, only 10 (30%) presented with isolated memory impairment. This figure is well within the range of previously reported rates. For example, Tabert and colleagues

found that, following comprehensive neuropsychological assessment, only 19% of their aMCI cases were categorised as pure aMCI,[5] whereas this figure reached 35% in the study by Alladi *et al*.[2] It should be borne in mind, however, that the rate of cases with purely amnestic MCI will vary in accordance with how impairment is defined. For example, Kramer and colleagues showed that the number of cases classified as pure aMCI was considerably higher (27%) when a cut-off of 1.5 s.d. below the mean, as opposed to 1 s.d. (resulting in a 5% rate), was used.[6] It therefore remains a possibility that our less stringent definition of impairment (i.e. 1 s.d. below the mean performance of the healthy control group) might have resulted in an overestimation of the frequency of cases with non-pure aMCI. Identifying accompanying non-memory cognitive impairment none the less appears important in light of recent evidence indicating a higher risk of conversion to Alzheimer's disease in patients with aMCI who show additional areas of cognitive impairment compared with patients with pure aMCI.[5]

The results of our study are also consistent with evidence indicating accompanying semantic memory impairment in aMCI,[1–5] with just 10 patients of 33 exhibiting episodic memory impairment in isolation, and 22 of the remaining 23 displaying evidence of semantic memory compromise. This finding may reflect an increased risk of conversion to Alzheimer's disease from aMCI, although early semantic memory failure is by no means specific to the former disease,[29,30] and although some studies report prognostic significance of performance on semantic memory measures,[18,31,32] others have failed to do so.[33] The stage at which impairments in this domain become apparent does appear to vary in accordance with the sensitivity of the measure employed.[4] The intact performance of participants with aMCI on measures of lexical (letter) fluency, also previously reported,[2,4] suggests that the 'initiation' aspects of semantic fluency tasks do not pose any difficulty to patients with this impairment subtype.

In view of the sound mean performances of our participants with aMCI on cognitive screening measures (MMSE score 28/30, Addenbrooke's Cognitive Examination score 88/100), it seems unlikely that consideration of such scores will be of any value in ruling out the presence of additional domains of cognitive impairment. Reliance on clinical judgement to determine the presence or absence of additional domains of subtly impaired cognition is similarly likely to prove difficult when dealing with patients with above-average premorbid IQ scores who are performing at sound levels on cognitive screens. Taken together, the above observations raise the question of whether global screening measures coupled with clinical judgement are a sufficient means of investigating MCI, and if not, whether additional resources or an expanded skill base will be required to handle this population clinically.

Our results reveal an absence of any significant difference in performance between the aMCI and control groups on measures of visuospatial function and processing speed. In-depth longitudinal evaluation of neuropsychological performance in MCI and questionable dementia suggests that visuospatial functions tend to fail secondarily to episodic memory and category fluency performances,[4,34] although some heterogeneity is known to exist.[35] It is therefore once again possible that our failure to demonstrate group differences on a visuospatial copying task reflects an earlier disease stage of our aMCI sample. Alternatively, it is conceivable that the varied and somewhat subjective scoring methods for the Rey Complex Figure Test copy task across different studies might be responsible for this finding. Cross-sectional findings pertaining to visuomotor processing speed in MCI vary, with some studies reporting significant differences between MCI and control groups[1,36,37] and others, like ourselves, failing to do so.[38–40] The disparity in findings may simply reflect the heterogeneity of aMCI or alternatively the disease stage. Group differences in processing speed might be more likely to exist where samples contain significant numbers of patients in the preclinical stages of a subcortical dementia of a cerebrovascular nature. For example, there is some evidence to suggest a disproportionately strong association between perceptual speed and parkinsonian signs in MCI.[41]

Limitations

Several study limitations should be noted. The significantly higher mean age of our patient group opens up the possibility that some of their performance deficits were explicable in terms of age-related cognitive decline. Ideally, control for age should have been better. However, aMCI group participants were identified on the basis of their performance on age-standardised tests; therefore, the discrepancy would not have influenced patient group membership. Longitudinal follow-up of these patients will help to clarify the relevance of this difference. Furthermore, our aMCI sample was characterised by a high average level of estimated premorbid general intellectual function, which introduces problems of applicability. Similar issues were present in a recent comparable study,[2] although other socio-demographic characteristics (i.e. gender, ethnicity, education and occupation) were not reported, preventing further comparison between that and our study. There may therefore be a need to replicate these findings employing greater numbers of age- and IQ-matched healthy controls and aMCI patients with average premorbid IQs, together with other socio-demographic markers more closely resembling the population mean.

Implications

Patients with MCI make up a significant number of referrals to older adult memory assessment services, with the most common referral subtype in our sample being that of aMCI, followed by equal numbers of non-amnestic MCI and worried well. Relatively few people with aMCI exhibit episodic memory compromise in isolation and fewer still show a visual but not verbal episodic memory deficit. Both the concept and criteria for MCI therefore appear to be relevant and indeed necessary adjuncts to clinical practice.

Our findings highlight the inherent difficulties of specifying a single measure in the assessment of memory and other cognitive functions in MCI, while at the same time emphasising the need for clarification of the means by which MCI criteria can be put into operation clinically. Initial attempts at better defining neuropsychological aspects of the aMCI criteria have been made,[1] but their application in a clinical sense remains inconsistent and their poor definition has not gone unnoticed.[42] The existence of a number of neuropsychological measures of well-documented sensitivity in aMCI and the strikingly similar mean performances of different clinic aMCI groups on such measures suggest that this need not be the case. Although the importance of exercising clinical judgement in arriving at a diagnosis of MCI cannot be ignored, it would none the less seem inevitable that further definition of the neuropsychological aspects of MCI criteria will be needed to facilitate identification of the subtypes of impairment and to further our understanding of their respective prognoses.

Acknowledgements

This study was financially supported by the Gordon Small Charitable Trust and the European Commission Network of Excellence Diagnostic Molecular Imaging (FP6-LIFESCI-

HEALTH Project Reference 512146). We thank all the participants and our clinical colleagues for referring their patients.

Jane A. Lonie, MSc, University of Edinburgh Division of Psychiatry, Royal Edinburgh Hospital, Edinburgh; **Lucie L. Herrmann**, MA, University of Oxford Department of Psychiatry, Warneford Hospital, Oxford; **Claire L. Donaghey**, MSc, University of Edinburgh Division of Psychiatry, Royal Edinburgh Hospital, Edinburgh; **Klaus P. Ebmeier**, FRCPsych, University of Edinburgh Division of Psychiatry, Royal Edinburgh Hospital, Edinburgh, and University of Oxford Department of Psychiatry, Warneford Hospital, Oxford, UK

Correspondence: Professor K. P. Ebmeier, Old Age Psychiatry Group, Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford OX3 7JX, UK. E-mail: klaus.ebmeier@psych.ox.ac.uk

First received 16 January 2007, final revision 11 April 2007, accepted 16 May 2007

References

1 Grundman M, Petersen RC, Ferris SH, Thomas RG, Aisen PS, Bennett DA, Foster NL, Jack CR, Jr, Galasko DR, Doody R, Kaye J, Sano M, Mohs R, Gauthier S, Kim HT, Jin S, Schultz AN, Schafer K, Mulnard R, van Dyck CH, Mintzer J, Zamrini EY, Cahn-Weiner D, Thal LJ. Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Arch Neurol* 2004; **61**: 59–66.

2 Alladi S, Arnold R, Mitchell J, Nestor PJ, Hodges JR. Mild cognitive impairment: applicability of research criteria in a memory clinic and characterization of cognitive profile. *Psychol Med* 2006; **36**: 507–15.

3 Ribeiro F, de Mendonça A, Guerreiro M. Mild cognitive impairment: deficits in cognitive domains other than memory. *Dement Geriatr Cogn Dis* 2006; **21**: 284–90.

4 Hodges JR, Erzinçlioglu S, Patterson K. Evolution of cognitive deficits and conversion to dementia in patients with mild cognitive impairment: a very-long-term follow-up study. *Dement Geriatr Cogn Dis* 2006; **21**: 380–91.

5 Tabert MH, Manly JJ, Liu X, Pelton GH, Rosenblum S, Jacobs M, Zamora D, Goodkind M, Bell K, Stern Y, Devanand DP. Neuropsychological prediction of conversion to Alzheimer disease in patients with mild cognitive impairment. *Arch Gen Psychiatry* 2006; **63**: 916–24.

6 Kramer JH, Nelson A, Johnson JK, Yaffea K, Glenna S, Rosena HJ, Millera BL. Multiple cognitive deficits in amnestic mild cognitive impairment. *Dement Geriatr Cogn Dis* 2006; **22**: 306–11.

7 Loewenstein DA, Acevedo A, Ownby R, Agron J, Barker WW, Isaacson R, Strauman S, Duara R. Using different memory cutoffs to assess mild cognitive impairment. *Am J Geriatr Psychiatry* 2006; **14**: 911–19.

8 Petersen RC, Stevens JC, Ganguli M, Tangalos EG, Cummings JL, DeKosky ST. Practice parameter: early detection of dementia: mild cognitive impairment (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 2001; **56**: 1133–42.

9 Petersen RC. Mild cognitive impairment as a diagnostic entity. *J Intern Med* 2004; **256**: 183–94.

10 Petersen RC, Morris JC. Mild cognitive impairment as a clinical entity and treatment target. *Arch Neurol* 2005; **62**: 1160–3.

11 Folstein MF, Folstein SE, McHugh PR. 'Mini-mental state': a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975; **12**: 189–98.

12 Mathuranath PS, Nestor PJ, Berrios GE, Rakowicz W, Hodges JR. A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. *Neurology* 2000; **55**: 1613–20.

13 Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* 1993; **43**: 2412–14.

14 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* (4th edn) (DSM–IV). APA, 1994.

15 Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey MB, Leirer VO. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res* 1982; **17**: 37–49.

16 Nelson H, Willison J. *National Adult Reading Test (NART).* nferNelson, 1991.

17 Brandt J. The Hopkins Verbal Learning Test: development of a new memory test with 6 equivalent forms. *Clin Neuropsychol* 1991; **5**: 125–42.

18 Blackwell AD, Sahakian BJ, Vesey R, Semple JM, Robbins TW, Hodges JR. Detecting dementia: novel neuropsychological markers of preclinical Alzheimer's disease. *Dement Geriatr Cogn Dis* 2004; **17**: 42–8.

19 Swainson R, Hodges JR, Galton CJ, Semple J, Michaele A, Dunna BD, Iddonf JL, Robbins TW, Sahakian BJ. Early detection and differential diagnosis of Alzheimer's disease and depression with neuropsychological tasks. *Dement Geriatr Cogn Dis* 2001; **12**: 265–80.

20 Rey A. Psychological examination of a case of post-traumatic encephalopathy. *Arch Psychol* 1941; **28**: 286–340.

21 McKenna P, Warrington EK. *Graded Naming Test.* nferNelson, 1983.

22 Thompson SA, Graham KS, Patterson K, Sahakian BJ, Hodges JR. Is knowledge of famous people disproportionately impaired in patients with early and questionable Alzheimer's disease? *Neuropsychology* 2002; **16**: 344–58.

23 Kaplan E, Goodglass H, Weintraub S. *Boston Naming Test.* Lea & Febiger, 1983.

24 Reitan RM. *Halstead–Reitan Neuropsychological Test Battery*. Reitan Neuropsychology Laboratory Press, 1985.

25 Lehrner J, Gufler R, Guttmann G, Maly J, Gleiss A, Auff E, Dal-Bianco P. Annual conversion to Alzheimer disease among patients with memory complaints attending an outpatient memory clinic: the influence of amnestic mild cognitive impairment and the predictive value of neuropsychological testing. *Wien Klin Wochenschr* 2005; **117**: 629–35.

26 Lopez OL, Jagust WJ, DeKosky ST, Becker JT, Fitzpatrick A, Dulberg C, Breitner J, Lyketsos C, Jones B, Kawas C, Carlson M, Kuller LH. Prevalence and classification of mild cognitive impairment in the Cardiovascular Health Study Cognition Study: Part 1. *Arch Neurol* 2003; **60**: 1385–9.

27 Lezak MD, Howieson DB, Loring DW, Hannay HJ, Fischer JS. *Neuropsychological Assessment*. Oxford University Press, 2004.

28 Collie A, Maruff P. The neuropsychology of preclinical Alzheimer's disease and mild cognitive impairment. *Neurosci Biobehav Rev* 2000; **24**: 365–74.

29 Clague F, Dudas RB, Thompson SA, Graham KS, Hodges JR. Multidimensional measures of person knowledge and spatial associative learning: can these be applied to the differentiation of Alzheimer's disease from frontotemporal and vascular dementia? *Neuropsychologia* 2005; **43**: 1338–50.

30 Graham NL, Emery T, Hodges JR. Distinctive cognitive profiles in Alzheimer's disease and subcortical vascular dementia. *J Neurol Neurosurg Psychiatr* 2004; **75**: 61–71.

31 Artero S, Tierney MC, Touchon J, Ritchie K. Prediction of transition from cognitive impairment to senile dementia: a prospective, longitudinal study. *Acta Psychiatr Scand* 2003; **107**: 390–3.

32 Estevez-Gonzàlez A, Garcia-Sànchez C, Boltes A, Otermin P, Pascual-Sedano B, Gironell A, Kulisevsky J. Semantic knowledge of famous people in mild cognitive impairment and progression to Alzheimer's disease. *Dement Geriatr Cogn Dis* 2004; **17**: 188–95.

33 Beatty WW, Salmon DP, Troster AI, Tivis RD. Do primary and supplementary measures of semantic memory predict cognitive decline by patients with Alzheimer's disease? *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn* 2002; **9**: 1–10.

34 Perry RJ, Hodges JR. Fate of patients with questionable (very mild) Alzheimer's disease: longitudinal profiles of individual subjects' decline. *Dement Geriatr Cogn Dis* 2000; **11**: 342–9.

35 Caine D, Hodges JR. Heterogeneity of semantic and visuospatial deficits in early Alzheimer's disease. *Neuropsychology* 2001; **15**: 155–64.

36 Arnaiz E, Almkvist O, Ivnik RJ, Tangalos EG, Wahlund LO, Winblad B, Petersen RC. Mild cognitive impairment: a cross-national comparison. *J Neurol Neurosurg Psychiatr* 2004; **75**: 1275–80.

37 Nordlund A, Rolstad S, Hellstrom P, Sjogren M, Hansen S, Wallin A. The Goteborg MCI study: mild cognitive impairment is a heterogeneous condition. *J Neurol Neurosurg Psychiatr* 2005; **76**: 1485–90.

38 Albert MS, Moss MB, Tanzi R, Jones K. Preclinical prediction of AD using neuropsychological tests. *J Int Neuropsychol Soc* 2001; **7**: 631–9.

39 Crowell TA, Luis CA, Rodney DV, Schinka JA, Mullan M. Memory patterns and executive functioning in mild cognitive impairment and Alzheimer's disease. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn* 2002; **9**: 288–97.

40 Fox NC, Warrington EK, Seiffer AL, Agnew SK, Rossor MN. Presymptomatic cognitive deficits in individuals at risk of familial Alzheimer's disease. A longitudinal prospective study. *Brain* 1998; **121**: 1631–9.

41 Boyle PA, Wilson RS, Aggarwal NT, Arvanitakis Z, Kelly J, Bienias JL, Bennett DA. Parkinsonian signs in subjects with mild cognitive impairment. *Neurology* 2005; **65**: 1901–6.

42 Portet F, Ousset PJ, Visser PJ, Frisoni GB, Nobili J, Scheltens Ph, Vellas B, Touchon J. Mild cognitive impairment (MCI) in medical practice: a critical review of the concept and new diagnostic procedure. Report of the MCI Working Group of the European Consortium on Alzheimer's Disease. *J Neurol Neurosurg Psychiatr* 2006; **77**: 714–18.

Psychological Medicine (2009), **39**, 23–31. © 2008 Cambridge University Press
doi:10.1017/S0033291708003346 Printed in the United Kingdom

ORIGINAL ARTICLE

Dual task performance in early Alzheimer's disease, amnestic mild cognitive impairment and depression

J. A. Lonie[1], K. M. Tierney[1], L. L. Herrmann[2], C. Donaghey[1], R. E. O'Carroll[3], A. Lee[1] and K. P. Ebmeier[1,2]*

[1] *University of Edinburgh Division of Psychiatry, The Royal Edinburgh Hospital, Edinburgh, UK*
[2] *University of Oxford Department of Psychiatry, Warneford Hospital, Oxford, UK*
[3] *University of Stirling Department of Psychology, University of Stirling, Stirling, UK*

Background. The dual task paradigm (Baddeley *et al*. 1986; Della Sala *et al*. 1995) has been proposed as a sensitive measure of Alzheimer's dementia, early in the disease process.

Method. We investigated this claim by administering the modified dual task paradigm (utilising a pencil-and-paper version of a tracking task) to 33 patients with amnestic mild cognitive impairment (aMCI) and 10 with very early Alzheimer's disease, as well as 21 healthy elderly subjects and 17 controls with depressive symptoms. All groups were closely matched for age and pre-morbid intellectual ability.

Results. There were no group differences in dual task performance, despite poor performance in episodic memory tests of the aMCI and early Alzheimer's disease groups. In contrast, the Alzheimer patients were specifically impaired in the trail-making test B, another commonly used test of divided attention.

Conclusions. The dual task paradigm lacks sensitivity for use in the early differential diagnosis of Alzheimer's disease.

Received 18 June 2007; Revised 12 December 2007; Accepted 28 February 2008; First published online 15 April 2008

Key words: Anterograde amnesia, depressive disorder, dysthymic disorder, geriatric assessment, memory disorders, neuropsychology.

Introduction

Alzheimer's disease (AD) is the most common form of dementia, estimated to rise dramatically in the future (Wimo *et al*. 2003). Research has focused on early accurate diagnosis and intervention. The construct 'amnestic mild cognitive impairment' (aMCI; Peterson *et al*. 2001) has become increasingly popular to predict those who are most at risk for developing dementia. It is considered a transitional stage between normal ageing and the earliest clinical diagnosis of AD (Petersen, 2005; Petersen & O'Brien, 2006). Research on clinic-based samples has suggested that the conversion rate from aMCI to dementia is 10–15% per year (e.g. Petersen *et al*. 1999; Storandt *et al*. 2006) compared with between 1% and 2% in a normal age-matched non-clinical sample.

While primary impairment in very early AD includes episodic memory function, many authors have reported that attention and executive functioning are also vulnerable at this stage (Parasuraman & Haxby, 1993; Perry & Hodges, 1999). In particular, people with early AD exhibit marked difficulty dividing their attention between two concurrent tasks. By comparing performance of a synchronous dual task with that of identical task components done separately and consecutively, a deficit in dual performance can be attributed to failure of the central executive that coordinates the simultaneous operation of these components (Baddeley *et al*. 1986). One advantage of the dual task paradigm is that it avoids modality-specific interference between tasks: the tracking task is presented visually and a manual response is required; information for the digit span task is presented aurally with a verbal response (Nebes *et al*. 2001). A further strength is that task demands can be fixed at individual ability levels, controlling for individual variation in performance in the component parts of the dual task. Therefore, each patient is his or her own control, adjusting for the generally poorer performance of AD patients in the baseline tasks (Logie *et al*. 2004).

Research has suggested that failure of the 'coordination' function is characteristic of mild AD in a laboratory setting. Participants with mild AD appear to

* Address for correspondence: Professor K. P. Ebmeier, University of Oxford, Section of Old Age Psychiatry, Warneford Hospital, Oxford OX3 7JX, UK.
(Email: Klaus.ebmeier@psych.ox.ac.uk)

be impaired, irrespective of task demands, and this impairment has been found to worsen with illness progression (Baddeley *et al*. 1986, 1991; MacPherson *et al*. 2004). Proponents of the dual task paradigm suggest such findings are in contrast to normal ageing, which they believe has a relatively minor effect on dual task performance (e.g. Baddeley *et al*. 1986; Hartley & Little, 1999; Logie *et al*. 2004; but see Crossley & Hiscock, 1992). The equipment used for this test is often an expensive computerized tracking device impractical for clinical settings (e.g. Baddeley *et al*. 1991; Logie *et al*. 2004). Della Sala *et al*. (1995) developed a modified pencil-and-paper version of the tracking component for the dual task. This has been reported to produce results comparable with the original instrument (Della Sala *et al*. 1995; Sebastian *et al*. 2006). To our knowledge the dual task paradigm has not been investigated with a sample defined according to recent aMCI criteria (Petersen *et al*. 1999).

The aim of this study was therefore to assess dual task performance in aMCI to ascertain whether this measure can be useful in the early diagnosis of AD. As AD is associated with a specific impairment in the aspect of working memory that coordinates performance of two separate tasks, we predicted that the performance of people with aMCI and very early AD should be significantly lower than that of aged matched controls. Furthermore, the inclusion of a group of elderly patients with symptoms of depression would test the specificity of dual task impairments in AD. On the basis of the previous research, we predicted that the depressed group would show impairment in the dual task compared with controls.

Method

Participants

We examined 33 patients with aMCI, 10 early AD patients, 17 control out-patients with depressive symptoms and 21 healthy elderly controls, following a protocol approved by the local ethics of research committee. All participants also took part in a larger longitudinal study of neuropsychological markers in pre-clinical AD. The aMCI patients were recruited over a 2-year period (September 2003–September 2005) from tertiary referrals to the local neuropsychological assessment service for older adults and met criteria for aMCI (Petersen *et al*. 1999). MCI patients had to give subjective reports of memory difficulty corroborated by an informant and exhibit objective memory impairments on neuropsychological tests of episodic memory. In terms of impairments on tests of episodic memory, 13 participants showed an impairment of more than 2 standard deviations (S.D.) below our control mean on two or more tests, a further four showed impairments of 1.5 to 2 S.D. on two or more tests, 12 participants were impaired at 1–1.5 S.D. below control means on two or more tests, and the final four participants performed more than 1 S.D. below controls on one episodic memory test. All aMCI patients underwent comprehensive neuropsychological and psychiatric evaluation and medical screening prior to study entry, as well as neuroimaging before or during the study period, if thought to be clinically indicated by the responsible specialist, i.e. in 24 of the 33 participants in this group. Exclusion criteria for the aMCI group were a diagnosis of dementia or other medical/neurological conditions which may account for memory loss, untreated depressive illness, significant or predominant cerebrovascular disease on neuroimaging, significant motor and/or visual problems or an age below 58 years. Mini Mental State Examination (MMSE) scores ranged from 24 to 30, with a mean of 28.3. The final aMCI group consisted of 15 males and 18 females with a mean age of 73.3 years (range 58–85 years).

For the healthy elderly control group (MMSE 28–30), we recruited spouses or carers of patients who had attended the service. Potential participants were excluded if there was a history of medical, psychiatric or neurological conditions (i.e. stroke or cerebrovascular disease, head injury, alcoholism, schizophrenia, etc) that could conceivably affect cognitive functioning. The healthy elderly control group was matched as closely as possible to the aMCI and early AD groups in terms of age and estimated pre-morbid intelligence quotient (IQ). The final elderly control group consisted of eight males and 13 females with had a mean age of 69.5 years (range 59–81 years).

Ten participants diagnosed with AD, in accordance with National Institute of Neurologic, Communicative Disorders and Stroke–AD and Related Disorders Association (NINCDS-ADRDA; McKhann *et al*. 1984) and DSM-IV (APA, 1994) diagnostic criteria, took part in the current study. AD patients were recruited from tertiary referrals to our neuropsychology service or via referrals to the local old age psychiatry service. All early AD patients scored above 23/30 on the MMSE and above 65/100 on the more comprehensive Addenbrooke's cognitive examination (ACE; Mathuranath *et al*. 2000), indicating relatively mild disease. Patients had undergone relevant medical screening and neuroimaging, together with comprehensive psychiatric and neuropsychological evaluation as part of their initial diagnostic workup. The final early AD group consisted of three males and seven females with a mean age of 73.6 years (range 65–81 years).

Seventeen participants with depressive symptoms (MMSE 25–30) were recruited via local psychiatric out-patient clinics and day hospitals. In an attempt to match this patient group with the aMCI group in terms of illness severity, patients with milder forms of depression were included. Fifteen of the 17 participants were receiving treatment for their symptoms at the time of testing; all but two of these pharmaceutical in nature. As it has been suggested that type of depression does not influence the magnitude of cognitive deficits (Christensen *et al*. 1997), participants with a variety of disorders were included. Eight patients had a history of major depression, two of bipolar disorder, two were suffering from anxiety disorders with depressive features, three were considered dysthymic and two were considered to be suffering with a subclinical level of depressive symptoms. Mean geriatric depression scale (30-item version) score for this group was 13.2 (range 0–27). We once again excluded patients with any medical, neurological or psychiatric condition with a known potential to affect cognitive function. The group consisted of three males and 14 females with a mean age of 73.3 years (range 65–84 years). Subjects gave informed written consent to the whole protocol which was approved by the Lothian Research Ethics Committee; the research was completed in accordance with the Helsinki Declaration.

Neuropsychological tests

All participants completed a variation on the modified dual task paradigm (Della Sala *et al*. 1995). This pencil-and-paper test of divided attention consists of two components (a digit span task and a visuospatial tracking task) that are each performed on their own before being performed concurrently. First, participants' digit span was determined. This involved repeating strings of digits read by an experimenter at a rate of approximately two per s. Initially, two-digit strings were presented and these increased one digit at a time if the participant correctly recited five of six examples of each length. When the participant failed to recite two or more strings of the same span, digit span for that person was considered to be the previous length. No time limit was imposed at this stage. Having determined the participants' individual digit span, participants had 90 s to recite as many digit strings, fixed at the individual participants' digit span, as possible (digit span – single). Responses were recorded as correct for each digit recited in the correct order.

Following this, participants completed the tracking task (Della Sala, 1999). An A3-sized sheet with 319 empty circles linked by a meandering line was presented to the participant. The participant was instructed to trace a line though circles, following the line that was already there, without lifting the pen from the paper. Participants had 90 s for this trial, and the number of circles reached during this time was recorded (tracking – single). The final trial was the concurrent dual task. Here participants had 90 s to simultaneously perform both tasks: recite digit strings fixed at their digit span (digit span – dual) as well as carrying out a tracking task identical to the one used above (tracking – dual). In order to take into account the various strategies one may adopt in performing the two tasks simultaneously, an overall decrement score was calculated using the following formula:

$$\mu = (1 - [(P_m + P_t)/2]) \times 100,$$

where μ is the combined dual task score, P_m is the proportional loss in span performance between single (X_{single}) and dual task (X_{dual}) conditions, $[(X_{single} - X_{dual})/X_{single}]$ while P_t is the equivalent proportional loss in tracking score. Thus a score of 100 would represent no dual task decrement and lower scores reflect greater dual task decrements.

A number of further tests were administered as part of the longitudinal investigation of neuropsychological markers. These included measures of general cognitive ability, such as the ACE, the more widely known MMSE and the National Adult Reading Test, revised version (NART-R; Nelson & Willison, 1991). The NART-R was used to provide an estimate of the pre-morbid level of intellectual functioning. Episodic memory was assessed using the Hopkin's verbal list test, revised (HVLT-R; Brandt, 1991) and the paired associates learning test (PAL) from the Cambridge automated neuropsychological test battery (Swainson *et al*. 2001). Participants also completed the trail-making test (TMT) part A and B (Reitan, 1985), considered a measure of attention and executive functioning.

The HVLT-R requires participants to recall as many words as possible immediately following presentation of a 12-item word list. The word list is presented on three consecutive learning trials. The participant is also required to recall, and finally recognize, as many words from the list as he or she is able, following a delay of 30 min. The PAL is a computerized measure of visuospatial learning requiring participants to learn the locations of an increasing number (i.e. 1, 2, 3, 6 and then 8) of patterns (Swainson *et al*. 2001). The score of interest was the number of pattern-position errors at the six pattern level. The TMT A requires tracing a line linking numbers in ascending order, while for the TMT B participants have to connect numbers and letters alternatively in ascending order: the

Table 1. *Demographic data*

Variable	Controls ($n=21$)	Depression ($n=17$)	aMCI ($n=33$)	Early AD ($n=10$)	Group differences
Males (n)	8	3	16	3	
Females (n)	13	14	17	7	
Age	69.5 (7.3)	73.3 (6.6)	73.1 (6.3)	73.6 (5.8)	–
NART	118.2 (2.9)	116.8 (6.2)	116.3 (8.5)	115.6 (5.5)	–
MMSE[a]	29.1 (0.7)	28.6 (1.5)	28.4 (1.6)	25.0 (2.3)	Controls = depression = aMCI > AD
ACE	94.6 (3.3)	91.7 (5.0)	89.0 (5.6)	76.7 (6.6)	Controls > aMCI > AD Depression > AD

aMCI, Amnestic mild cognitive impairment; AD, Alzheimer's disease; NART, National Adult Reading Test; MMSE, Mini Mental State Examination; ACE, Addenbrooke's cognitive examination.

Values are given as mean (standard deviation).

[a] Games–Howell multiple comparison carried out due to lack of homogeneity of variances.

participant has to divide his/her attention back and forth between multiple lines of thought.

Each of these measures has been shown to be sensitive to very early AD (Chen *et al*. 2000; Nathan *et al*. 2001; Swainson *et al*. 2001; Hogervorst *et al*. 2002; Blackwell *et al*. 2004; Stokholm *et al*. 2006). Neuropsychological assessments lasted approximately 90 min in total. The order of test administration was identical for all assessments.

Statistics

Data were analysed using SPSS 12.0 for Windows (SPSS Inc., Chicago, IL, USA). Demographic variables were analysed using univariate analysis of variance (ANOVA), and Tukey honestly significantly different pairwise comparisons were carried out on all significant analyses where possible. Where the assumption of homogeneity of variance was not met, this was adjusted for using Games–Howell *post-hoc* pairwise comparisons, given that the sample sizes were unequal in the current analysis. A univariate ANOVA was carried out on the overall decrement score (see above). Decrement scores broken down into tracking decrement and digit span decrement were also calculated and examined using ANOVAs. Two participants in the early AD group were incapable of completing the TMT B; in these cases a default ceiling score of 500 s to completion was applied.

Results

Participant characteristics

Demographic matching characteristics are presented in Table 1. There were no group differences in age [$F(3,77)=1.73$] or estimated pre-morbid full-scale IQ [$F(3,75)=0.55$]. The mean MMSE score for the early AD group was, as expected, significantly lower than that of the other groups [$F(3,77)=17.70$, $p<0.0001$] (AD *v*. healthy controls, $p=0.001$; AD *v*. controls with depressive symptoms, $p<0.005$; AD *v*. controls, $p<0.005$). No other group differences in mean MMSE score were noted. As expected, the early AD patients had significantly lower mean ACE scores than did all other groups [$F(3,77)=29.30$, $p<0.0001$] (*post-hoc* tests as above in all cases, $p<0.0001$). The ACE also discriminated between normal elderly control participants and aMCI patients, with the latter group obtaining a significantly lower mean ACE score (*post hoc* $p=0.01$).

Dual task performance

Group means and S.D.s for the digit span task and the tracking measures of the modified dual task paradigm are presented in Table 2. Mean percentage scores for performance on the concurrent tasks, the digit span tasks and the visuospatial tracking tasks for each of the four groups are presented in Table 3. On carrying out a one-way non-repeated ANOVA on the overall decrement score, no group difference was found [$F(3,77)=0.63$]. Similarly, no significant group differences were found for any of the other component tasks or decrement scores.

Other cognitive functions

Group mean scores and S.D.s for the HVLT-R, the number of errors at the six pattern level of the PAL and the TMT B are presented in Table 4. On analysing the HVLT-R delayed recall data, there was a significant group effect [$F(3,77)=12.39$, $p<0.0001$]. On closer analysis, the AD group recalled significantly fewer words than the healthy control ($p<0.0001$) and depression groups ($p<0.0001$). Similarly, the aMCI

Table 2. *Digit span and individual component measures of the dual task (span and tracking, performed separately and together)*

Task	Controls ($n=21$)	Depression ($n=17$)	aMCI ($n=33$)	Early AD ($n=10$)
Digit span	5.5 (0.7)	5.8 (1.0)	5.6 (0.9)	5.1 (0.7)
Digit span (single)[a]	1.0 (0.03)	0.9 (0.05)	1.0 (0.05)	1.0 (0.03)
Digit span (dual)[a]	0.9 (0.05)	0.9 (0.08)	0.9 (0.08)	1.0 (0.02)
Tracking (single)[b]	141 (56.5)	140 (51.7)	126 (38.9)	120 (46.3)
Tracking (dual)[b]	122 (46.0)	126 (58.3)	114 (36.3)	107 (35.6)

aMCI, Amnestic mild cognitive impairment; AD, Alzheimer's disease.
Values are given as mean (standard deviation).
[a] Proportion of digits recalled in the correct position (1 = all correct).
[b] Number of circles joined in 90 s.

Table 3. *Percentage loss of performance in component tasks and overall decrement score during the dual task*[a]

Task	Controls ($n=21$)	Depression ($n=17$)	aMCI ($n=33$)	Early AD ($n=10$)
Digit span	96 (3.8)	97 (8.6)	98 (7.6)	100 (3.3)
Tracking	90 (22.8)	88 (16.8)	92 (15.4)	93 (17.6)
Overall decrement	93 (11.1)	92 (8.2)	95 (8.6)	97 (9.1)

aMCI, Amnestic mild cognitive impairment; AD, Alzheimer's disease.
Values are given as mean (standard deviation).
[a] Percentage loss of performance scores were calculated as $(1-[(X_{single}-X_{dual})/X_{single}])\times 100$ and the overall decrement score as $\mu=(1-[(P_m+P_t)/2])\times 100$, as described in the Method section.

Table 4. *Other cognitive domain measures*

Task	Controls ($n=21$)	Depression ($n=17$)	aMCI ($n=33$)	Early AD ($n=10$)	Group differences
HVLT-R delay	8.1 (2.8)	8.1 (3.3)	4.9 (3.3)	2.1 (3.7)	Controls = depression > aMCI = AD
PAL errors[a]	7.8 (6.9)	10.9 (7.8)	16.5 (12.9)	40.7 (10.6)	Controls, depression, aMCI < AD Controls < aMCI
TMT A	40.3 (11.2)	54.1 (23.1)	49.6 (36.1)	57.6 (25.3)	–
TMT B	87.6 (31.5)	134.2 (53.6)	106.3 (49.4)	216.7 (157.7)	Controls < depression Controls, depression, aMCI < AD[b]

aMCI, Amnestic mild cognitive impairment; AD, Alzheimer's disease; HVLT-R, Hopkins verbal learning test, revised; PAL errors, six pattern stage errors from the paired associates learning test; TMT A, trail-making test part A; TMT B, trail-making test part B.
Values are given as mean (standard deviation).
[a] Games–Howell multiple comparison was used because of unequal variances.
[b] After removing effects of TMT A (see text).

group performed more poorly than the healthy control ($p<0.005$) and depression groups ($p<0.01$). No significant difference was found between the AD and aMCI groups. The performance of the elderly control and depression groups on the HVLT-R delayed recall did not differ. However, the AD group made significantly more errors at the six pattern stage of the PAL compared with all other groups [$F(3,755)=22.82$,

$p < 0.0001$] (*post hoc* tests comparing AD with other groups were in all cases $p < 0.0001$). The aMCI group's error scores fell between those of the healthy control and AD groups, and significantly differed from both of these (aMCI *v.* healthy controls, $p < 0.05$; aMCI *v.* AD, $p < 0.0001$). A significant group effect was also found for the TMT B [$F(3,77) = 8.62$, $p < 0.0001$]. In the *post hoc* analyses, only the control and depression groups differed in terms of TMT B scores ($p < 0.05$); participants with depressive symptoms took significantly longer to complete the task. However, once time to completion on TMT part A (a measure of psychomotor speed) was statistically controlled for, a different pattern of group differences emerged [$F(3,76) = 7.76$, $p < 0.0001$]. Specifically, it was found that the participants with AD took longer to complete the second task compared with all other groups (aMCI *v.* AD, $p < 0.0001$; healthy controls *v.* AD, $p < 0.0001$; controls with depressive symptoms *v.* AD, $p < 0.05$). The group difference between the control and depressive symptom groups was no longer significant. No other group differences were uncovered.

Discussion

This study investigated the claim that the dual task paradigm can be used in the early diagnosis of dementia of the Alzheimer's type. We assessed the concurrent performance of a visuospatial tracing task and a digit span forward task in four diagnostic groups with aMCI (MMSE 24–30), early AD (MMSE 23–29), symptoms of depression (MMSE 25–30) and healthy elderly controls (MMSE 28–30). Our results show that aMCI is not associated with impaired dual task performance; those with aMCI had comparable performance to healthy older adults and older adults with depressive symptoms. Our early AD group was similarly unimpaired on the modified dual task paradigm relative to depressive and non-depressive elderly control groups and the presence of depressive symptoms appeared to have no effect on dual task performance. By contrast, and indeed by definition, episodic memory impairments were present in the aMCI and early AD groups. The early AD group also exhibited an impaired ability to divide their attention at pace, as indicated by part B of the TMT.

These results shed some light on previous findings. One line of research has suggested that dual task performance is vulnerable to the influence of AD, even early in the disease course (Baddeley *et al*. 2001; Logie *et al*. 2004). However, such studies generally involve participants varying in severity from minimal to mild AD. When participants with AD are divided by severity using the MMSE, only the more severely ill patients (e.g. MMSE < 24) are impaired on the dual task paradigm (Greene *et al*. 1995; Perry *et al*. 2000; Crossley *et al*. 2004). This result is in agreement with the absence of impairment on the dual task measure observed in the current study in early AD. The combined findings suggest that dual task impairments are generally not observed early on in the AD process, with MMSE scores above 23/30.

Only one other study has investigated the dual task performance of a group of older adults with cognitive impairment without a diagnosis of dementia (Holtzer *et al*. 2004). Cognitively impaired adults, defined by a dementia rating scale (DRS) cut-off score of < 124 (Mattis, 1988), performed two tasks in different modalities at the same time. Two combinations of tests were used: a visual cancellation task (where participants were required to cross out a specified stimulus type from a field of stimuli) combined with a digit span task, and the same visual cancellation task combined with a verbal fluency task. The researchers report that their cognitively impaired group exhibited a significantly larger dual task decrement than age-matched controls. However, the cognitively impaired group in the Holtzer *et al*. (2004) study was identified solely on the basis of a DRS cut-off score falling at or below levels that are indicative of an underlying dementia. It is for this reason difficult to be certain of, or to compare, disease severity of this 'minimally cognitively impaired' group with other studies, which commonly use well-established clinical and research criteria to define patient groups. Furthermore, the cognitively impaired group in the Holtzer *et al*. (2004) study were significantly less well educated than the control groups, while in the current study participant groups were well matched both in terms of age and estimated levels of pre-morbid intelligence.

Holtzer *et al*. (2004) did not investigate the potential influence of depression on dual task performance. This is crucial where consideration is being given to the early and differential diagnostic value of a neuropsychological measure. Hasher & Zacks (1979) confirmed our result that people with depression show impaired attention during effortful processing tasks, for instance on measures of divided attention such as the TMT B (Nathan *et al*. 2001; Mahurin *et al*. 2006). Only one study has investigated the effect of depressive symptoms on Baddeley *et al*.'s (1986) original dual task paradigm (Nebes *et al*. 2001). This indicated that people with depression had a significantly greater decrement in computerized tracking performance and a composite decrement measure than non-depressed controls. No study to date has investigated the effects of clinically depressed mood on the modified version of the dual task paradigm to replicate or contradict our negative result (Della Sala *et al*. 1995).

A strength of the current investigation relates to the availability of additional neuropsychological data demonstrating the existence of significant episodic memory impairments in aMCI and early AD and additional impairment of speeded divided attention (as assessed by TMT B) in early AD. The TMT B assesses the ability to divide attention back and forth between multiple lines of thought (connecting numbers and letters, respectively), but differs from the dual task paradigm in that its different components are not drawn from separate modalities. Performance is thus more vulnerable to reduced processing capacity. Several previous studies have demonstrated that TMT B is impaired in the very early and even pre-clinical stages of AD (Lafleche & Albert, 1995; Arnaiz *et al*. 2000; Perry *et al*. 2000; Nathan *et al*. 2001; Crowell *et al*. 2002; Crossley *et al*. 2004; Alladi *et al*. 2006; Baudic *et al*. 2006; Stokholm *et al*. 2006), although its specificity for AD, as distinct from, for example, depression, has not been established.

The Holtzer *et al*. (2004) study compared the dual task performance of minimally cognitively impaired participants only with their performance on tests comprising the single task conditions (i.e. visual cancellation, digit span and letter fluency). However, these tests are not, generally speaking, associated with impairments in very early and pre-clinical AD and it is therefore not surprising that they are insensitive to cognitive deficits in the minimally impaired group, as was the case in this study.

One important methodological feature may have influenced the current results: While those studies reporting general dual task impairment in early AD used both computerized and pencil-and-paper versions of the tracking task, only the modified version utilising the pencil-and-paper tracking task (Della Sala *et al*. 1995) has been used in studies that separated participants by symptom severity. Thus, while patients who are minimally affected do not show impairments on the modified version of the task, it remains possible that they would show impairments if the test were more taxing – for instance if the dual task paradigm included the original computerized version of the tracking task. This version of the task requires increased effort and attention, as participants are required to adjust to an external influence (i.e. the speed of the light dot on the screen) rather than working at a self-defined rate. It may therefore be sufficiently taxing to identify those who are not picked up by the more straightforward pencil-and-paper tracking task. However, the paper-and-pencil version (as opposed to the computerized task) is more likely to be adopted for widespread use in clinical and research practice, which underscores the relevance of our negative result.

A further methodological issue is the variability of dual task administration, which can lead to difficulties comparing findings across studies. We administered each of the three trials in blocks of 90 s, while some previous studies set the trial time at 120 s (e.g. Perry *et al*. 2000). Most dual task studies have utilized pencil-and-paper tracking tasks that required participants to cross out boxes on an A4-size sheet of paper to form a chain (e.g. Baddeley *et al*. 1997). The current task required participants to trace a line through linked empty circles on an A3-size sheet. While the initial dual task paradigm involved recording the number of completely correct digit strings (Baddeley *et al*. 1986), many subsequent studies, including the current investigation, have calculated the number of digits recalled in the correct order for this measure. The significance of such alterations to dual task administration requires further investigation.

A partial alternative explanation for our negative result is that a majority of individuals forming our aMCI group may fail to convert to AD in the future. If this proves to be the case, then the absence of dual task impairment would not be surprising. The issue will be resolved through the longitudinal follow-up of participants with aMCI, currently underway. However, the sound performance of our early AD group on the dual task measure makes it more likely that the negative result for our aMCI patients is due to lack of test sensitivity rather than absence of underlying AD pathology. The impaired performance of the early AD group on an alternative popular measure of speeded divided attention implies that the dual task measure lacks sensitivity to very early changes of an attentional/executive nature in AD.

In conclusion, people with early AD and aMCI did not display impaired performance on the modified version of the dual task paradigm at a time when episodic memory, and in the case of early AD, speeded divided attention, were significantly impaired. The likely explanation is that the dual task paradigm is insufficiently sensitive for use as an adjunctive cognitive tool in the early diagnosis of AD. Future longitudinal research is needed to investigate the use of dual task tests of varying demand in aMCI and very early AD participants in an effort to determine the potential influence of task demands and complexity on performance.

Acknowledgements

The authors are grateful to the Gordon Small Charitable Trust for financial support and to clinical colleagues for referring their patients.

Declaration of Interest

None.

References

Alladi S, Arnold R, Mitchell J, Nestor PJ, Hodges JR (2006). Mild cognitive impairment: applicability of research criteria in a memory clinic and characterization of cognitive profile. *Psychological Medicine* **36**, 507–513.

APA (1994). *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association: Washington, DC.

Arnaiz E, Blomberg M, Fernaeus SE, Wahlund LO, Winbald B, Almkvist O (2000). Psychometric discrimination of Alzheimer's disease and mild cognitive impairment. *Alzheimer's Reports* **2**, 97–103.

Baddeley AD, Baddeley HA, Bucks RS, Wilcock GK (2001). Attentional control in Alzheimer's disease. *Brain* **124**, 1492–1508.

Baddeley AD, Bressi S, Della Sala S, Logie R, Spinnler H (1991). The decline of working memory in Alzheimer's disease: a longitudinal study. *Brain* **114**, 2521–2542.

Baddeley AD, Della Sala S, Papagno C, Spinnler H (1997). Dual-task performance in dysexecutive and nondysexecutive patients with a frontal lesion. *Neuropsychology* **11**, 187–194.

Baddeley AD, Logie R, Bressi S, Della Sala S, Spinnler H (1986). Dementia and working memory. *Quarterly Journal of Experimental Psychology Section A Human Experimental Psychology* **38**, 603–618.

Baudic S, Barba GD, Thibaudet MC, Smagghe A, Remy P, Traykov L (2006). Executive function deficits in early Alzheimer's disease and their relations with episodic memory. *Archives of Clinical Neuropsychology* **21**, 15–21.

Blackwell AD, Sahakian BJ, Vesey R, Semple JM, Robbins TW, Hodges JR (2004). Detecting dementia: novel neuropsychological markers of preclinical Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders* **17**, 42–48.

Brandt J (1991). The Hopkins verbal learning test: development of a new memory test with 6 equivalent forms. *Clinical Neuropsychologist* **5**, 125–142.

Chen P, Ratcliff G, Belle SH, Cauley JA, DeKosky ST, Ganguli M (2000). Cognitive tests that best discriminate between presymptomatic AD and those who remain nondemented. *Neurology* **55**, 1847–1853.

Christensen H, Griffiths K, Mackinnon A, Jacomb P (1997). A quantitative review of cognitive deficits in depression and Alzheimer-type dementia. *Journal of the International Neuropsychological Society* **3**, 631–651.

Crossley M, Hiscock M (1992) Age-related differences in concurrent-task performance of normal adults: evidence for a decline in processing resources. *Psychology and Aging* **7**, 499–506.

Crossley M, Hiscock M, Foreman JB (2004). Dual-task performance in early stage dementia: differential effects for automatized and effortful processing. *Journal of Clinical and Experimental Neuropsychology* **26**, 332–346.

Crowell TA, Luis CA, Vanderploeg RD, Schinka JA, Mullan M (2002). Memory patterns and executive functioning in mild impairment and Alzheimer's disease. *Aging Neuropsychology and Cognition* **9**, 288–297.

Della Sala S (1999). Paper and Pencil Dual Task – University of Edinburgh (http://www.psy.ed.ac.uk/people/sdsala/tests/sdsdualtask). Accessed 1 February 2008.

Della Sala S, Baddeley A, Papagno C, Spinnler H (1995). Dual task paradigm: a means to examine the central executive. In *Annals of the New York Academy of Sciences, vol. 769. Structure and Functions of the Human Prefrontal Cortex* (ed. J. Graham, K. J. Holyoak and F. Boller), pp. 161–171. New York Academy of Sciences: New York.

Greene JDW, Hodges JR, Baddeley AD (1995). Autobiographical memory and executive function in early dementia of Alzheimer type. *Neuropsychologia* **33**, 1647–1670.

Hartley AA, Little DM (1999). Age related differences and similarities in dual-task interference. *Journal of Experimental Psychology: General* **128**, 416–449.

Hasher L, Zacks RT (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General* **108**, 356–388.

Hogervorst E, Combrinck M, Lapuerta P, Rue J, Swales K, Budge M (2002). The Hopkins verbal learning testing and screening for dementia. *Dementia and Geriatric Cognitive Disorders* **13**, 13–20.

Holtzer R, Burright RG, Donovick PJ (2004). The sensitivity of dual-task performance to cognitive status in aging. *Journal of the International Neuropsychological Society* **10**, 230–238.

Lafleche G, Albert MS (1995). Executive function deficits in mild Alzheimer's disease. *Neuropsychology* **9**, 313–320.

Logie RH, Cocchini G, Della Sala S, Baddeley AD (2004). Is there a specific executive capacity for dual task coordination? Evidence from Alzheimer's disease. *Neuropsychology* **18**, 504–513.

MacPherson SE, Della Sala S, Logie RH (2004). Dual-task interference of encoding and retrieval processes in healthy and impaired working memory. *Cortex* **40**, 183–184.

Mahurin RK, Velligan DI, Hazleton B, Davis JM, Eckert S, Miller AL (2006). Trail making test errors and executive function in schizophrenia and depression. *Clinical Neuropsychologist* **20**, 271–288.

Mathuranath PS, Nestor PJ, Berrios GE, Rakowicz W, Hodges JR (2000). A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. *Neurology* **55**, 1613–1620.

Mattis S (1988). *Dementia Rating Scale* (*DRS*). Psychological Assessment Resources, Inc.: Odessa, FL.

McKhann G, Drachman D, Folstein M, Katzmann R, Price D, Stadlan EM (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRADA work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology* **34**, 939–944.

Nathan J, Wilkinson D, Stammers S, Low JL (2001). The role of tests of frontal executive functioning in the detection of mild dementia. *International Journal of Geriatric Psychiatry* **16**, 18–26.

Nebes RD, Butters MA, Houck PR, Zmuda MD, Aizenstein H, Pollock BG, Muslant BH, Reynolds CF (2001). Dual-task performance in depressed geriatric patients. *Psychiatry Research* **102**, 139–151.

Nelson HE, Willison JR (1991). *The Revised National Adult Reading Test – Test Manual*. NFER-Nelson: Windsor, UK.

Parasuraman R, Haxby J (1993). Attention and brain function in Alzheimer's disease: a review. *Neuropsychology* **7**, 242–272.

Perry RJ, Hodges JR (1999). Attention and executive deficits in Alzheimer's disease: a critical review. *Brain* **122**, 383–404.

Perry RJ, Watson P, Hodges JR (2000). The nature and staging of attention dysfunction in early (minimal and mild) Alzheimer's disease: relationship to episodic and semantic memory impairment. *Neuropsychologia* **38**, 252–271.

Petersen RC (2005). Mild cognitive impairment as a clinical entity and treatment target. *Archives of Neurology* **62**, 1160–1163.

Petersen RC, Doody R, Kurz A, Morris JC, Rabins JC, Ritchie K, Rossor M, Thal L, Winbald B (2001). Current concepts in mild cognitive impairment. *Archives of Neurology* **58**, 1985–1992.

Petersen RC, O'Brien J (2006). Mild cognitive impairment should be considered for DSM-V. *Journal of Geriatric Psychiatry and Neurology*, **19**, 147–154.

Petersen RC, Smith GE, Waring SC, Ivrnik RJ, Kokmen E (1999). Mild cognitive impairment: clinical characterization and outcome. *Archives of Neurology* **56**, 303–308.

Reitan RM (1985). *Halstead–Reitan Neuropsychological Test Battery*. Reitan Neuropsychology Laboratory Press: Tucson, AZ.

Sebastian MV, Menor J, Elousa MR (2006). Attentional dysfunction of the central executive in AD: evidence from dual task and perseveration errors. *Cortex* **42**, 1015–1020.

Stokholm J, Vogel A, Gade A, Waldemar G (2006). Heterogeneity in executive impairment in patients with very mild Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders* **22**, 54–59.

Storandt M, Grant EA, Miller JP, Morris JC (2006). Longitudinal course and neuropathologic outcomes in original vs. revised MCI and in pre-MCI. *Neurology* **67**, 467–473.

Swainson R, Hodges JR, Galton CJ, Semple J, Michael A, Dunn BD, Iddon JL, Robbins TW, Sahakian BJ (2001). Early detection and differential diagnosis of AD and depression with neuropsychological tasks. *Dementia and Geriatric Cognitive Disorders* **12**, 265–280.

Wimo A, Winbald B, Aguero-Torres H, von Strauss E (2003). The magnitude of dementia occurrence in the world. *Alzheimer's Disease and Associated Disorders* **17**, 63–67.

MIND & BRAIN, THE JOURNAL OF PSYCHIATRY REVIEW ARTICLE

A Systematic Review of Cognitive Screening for Mild Cognitive Impairment

Jane A Lonie[1], *Ukwuori G Kalu*[2] *and Klaus P Ebmeier*[1,2]

Affiliations: [1]*Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK and* [2]*Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK*

ABSTRACT

BACKGROUND:

Although patients with mild cognitive impairment (MCI) make up a significant proportion of outpatient and memory clinic visits in old age psychiatry, their cognitive evaluation is typically restricted to brief dementia screens. We present an evaluation of published cognitive screens in MCI, their validity, and utility.

METHOD:

Papers published after Petersen's original MCI criteria were identified with 10 library search engines, combining the search terms ''mild cognitive impairment'' and ''cognitive screening'', as well as ''mild cognitive impairment'' and the names of 39 screening tests recently identified in a relevant review.

RESULTS:

Twenty-six relevant publications were identified, of which seven summarized the sensitivity and specificity of four comprehensive screening tests, and 19 evaluated 11 non-comprehensive tests. A recently published meta-analysis of the use of the mini-mental state examination summarized five mild cognitive impairment studies [3]. Sensitivities over 80% for detecting MCI among healthy volunteers were reported for all four comprehensive and three of the 11 non-comprehensive screening tests. Equivalent specificity values over 80% could be found for two comprehensive and five non-comprehensive tests. With the exception of six studies, MCI study sample sizes were universally small (i.e., $n \leq 100$), and prognostic values were reported for only two of the identified 15 screening measures.

CONCLUSIONS:

Sensitivities of the full domain measures were universally high, but information about their specificity against psychiatric and non-progressive neurological conditions and predictive validity is largely lacking. Conversely, non-comprehensive screening tests appear to be less accurate overall, with relatively greater specificity and lesser sensitivity.

1

Keywords: systematic review, screening, mild cognitive impairment, dementia, Alzheimer's disease, differential diagnosis, neuropsychology

Correspondence: KP Ebmeier, Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford, OX3 7JX, UK. e-mail: Klaus.ebmeier@psych.ox.ac.uk

BACKGROUND

Mild cognitive impairment (MCI) is used to describe cognitive impairment in patients who do not fulfill the criteria for dementia. The presenting complaint is usually recent memory loss. Amnestic MCI (aMCI) may be differentiated from early dementia on the basis of sound functional capacity, a non-progressive course, or intact functioning in other ''non-memory'' cognitive domains. Although at elevated risk of developing dementia, not all aMCI patients do so. Cognitive impairment in aMCI patients attending memory clinics is, however, usually seen to persist or gradually worsen, with a mean conversion rate to dementia of 10% per annum [1, 4–10].

Cognitive impairment is common, and patients with aMCI make up about one fifth of referrals to specialist memory clinics [7, 11]. Cognitive evaluation of aMCI, therefore, forms an important part of geriatric care. There are no consensus guidelines as to how the cognitive aspects of aMCI should be assessed [12]. At present, clinical cognitive examinations are largely restricted to brief screening measures, the vast majority of which were designed to evaluate early- and moderate-stage dementia. There is consequently a need for up-to-date information about how these measures ''perform'' as screening tools for aMCI.

The utility of a cognitive screening measure may be judged on a number of levels. The primary role of cognitive screening is to detect cognitive impairment. The screening measure must be capable of picking up clinically meaningful levels of cognitive impairment where these exist in aMCI and early-stage dementia. The issue of *sensitivity* is therefore pertinent.

Mind & Brain, the Journal of Psychiatry jop218262.3d 5/3/10 19:50:03
The Charlesworth Group, Wakefield +44(0)1924 369598 - Rev 9.0.225/W (Oct 13 2006)

Conversely, cognitive decline(s) is also observed in association with the normal aging processes and a wide variety of neurological, psychiatric, and medical illnesses among the elderly. Screening instruments therefore need to be *specific* against cognitive changes of a benign or long-standing nature.

On account of the slowly progressive nature of neurodegenerative illnesses, particularly in their early and preclinical phases, it is often necessary to readminister a cognitive screening instrument in order to determine the course of cognitive impairment, i.e., stable or progressive. For this reason, the stability of a cognitive screening measure across both time and raters is important, as those instruments associated with lesser degrees of random variation afford the clinician the opportunity to detect smaller amounts of meaningful cognitive decline.

In screening for cognitive impairment, there is an inevitable tradeoff between the ''time constraints'' of the clinician and the ''breadth of information'' that is obtainable. Ideally, a cognitive screening instrument should be *brief* and *user friendly* enough to facilitate usage in a wide range of care settings, including that of primary care. The availability of *cutoff values* for ease of interpretation is particularly useful in this regard. Increasingly, however, more in-depth cognitive evaluation is recognized as a key component in the early and differential diagnostic process for dementia. At more specialist levels of care, representation of each of the *primary domains of cognition* (i.e., memory, language, visuo-spatial and perceptual function, attention, and executive function) is crucial, but of course associated with longer test administration times.

Here, we review the utility of cognitive screening instruments in aMCI with reference to sensitivity to aMCI and specificity against healthy control subjects (and, where information is available, other neurodegenerative and non-neurodegenerative conditions that also give rise to cognitive impairment). Test–retest reliability across both time and different raters, administration times, availability and ease of use of normative data, and the degree of cognitive domain coverage (where ''comprehensive'' equates to coverage of each of the primary domains of cognition and ''non-comprehensive'' equates to incomplete coverage) are also considered.

METHODS

Two literature searches were conducted replicating our earlier method, and the results merged with the data already published [13]. Both searches were restricted to papers published after the formulation of Petersen *et al*'s initial set of criteria for MCI in 1999, in order to ensure a reasonable level of cross-study homogeneity [1, 14]. In the first, the search terms ''mild cognitive impairment'' and ''cognitive screening'' were entered into the following electronic databases: BIOS Previews, ISI Web of Knowledge, ISI Proceedings, Embase, MedLine, PsychInfo, SSCI, ASSIA Plus, Psych Articles. In the second search, the term ''mild cognitive impairment'' was entered into each of the above databases in combination with the names of 39 screening instruments identified by Cullen *et al* [2] in a recent review of tests for cognitive impairment.

The results of the two literature searches were combined with our previously published database [13]. Duplicate and repeat publications were removed alongside studies not in English and those reporting findings for screening instruments of a non-cognitive or non-face-to-face nature (i.e., olfactory, telephone, or internet administered or wholly carer- or informant-rated instruments). In the interests of creating a division between ''cognitive screening'' instruments and more formal and standardized ''neuropsychological'' measures, we also excluded those studies in which the measure of interest was of a non-screening nature (i.e., where this comprised one or more well-known standardized neuropsychological measures, not commonly utilized for screening). If patients' data had been published repeatedly, the most recent or inclusive publication was chosen. Data were extracted systematically and tabulated.

RESULTS

The combined searches yielded a total of 94 papers, 19 of which were excluded as they comprised duplicated data. Of the remaining 75 papers, a further 16 were excluded as they were not face to face ($n=10$), non-cognitive in nature ($n=4$), or the findings were reported in a language other than English ($n=2$). None of the screening measures reported in the non-English papers were reviewed as part of the included papers. Further studies were excluded as they did not meet the study objectives, i.e., ''to investigate the screening utility of cognitive screening measures in aMCI'' ($n=32$), where Petersen *et al*'s [1] aMCI criteria were not met or specified ($n=7$), where no full text version was available ($n=2$), as well as books ($n=1$) and abstracts ($n=16$).

A total of three new papers were therefore retained [15–17]. To these, a further 23 were added from Lonie *et al* [13], giving a total of 26 papers comprising the focus of this review. A recent meta-analysis of the diagnostic accuracy of the mini-mental state examination in aMCI was also included [3] (**Figure 1**). 2

Comprehensive cognitive screening

Table 1 summarizes studies included in the review, divided into comprehensive and non-comprehensive screening tests. 3 All four comprehensive screening tests achieved sensitivities over 80% for detecting MCI among healthy volunteers, but equivalent specificity values could only be found for two comprehensive tests (ACE-R [29]; MoCA [16, 41]). All MCI study sample sizes were small (i.e., $n \leq 100$), and prognostic values were reported for only one of the four identified comprehensive screening measures (CAMCOG [28]). None of the measures report specificity values for MCI in relation to depression or any other neurological or psychiatric diseases of a non-progressive nature. One-month retest reliability values of 70–92% have been reported for the MoCA [41], but not for follow-up over clinically relevant intervals (6–12 months), or for any other of the comprehensive cognitive screening measures. None of the comprehensive tests has

Mind & Brain, the Journal of Psychiatry jop218262.3d 5/3/10 19:50:04
The Charlesworth Group, Wakefield +44(0)1924 369598 - Rev 9.0.225/W (Oct 13 2006)

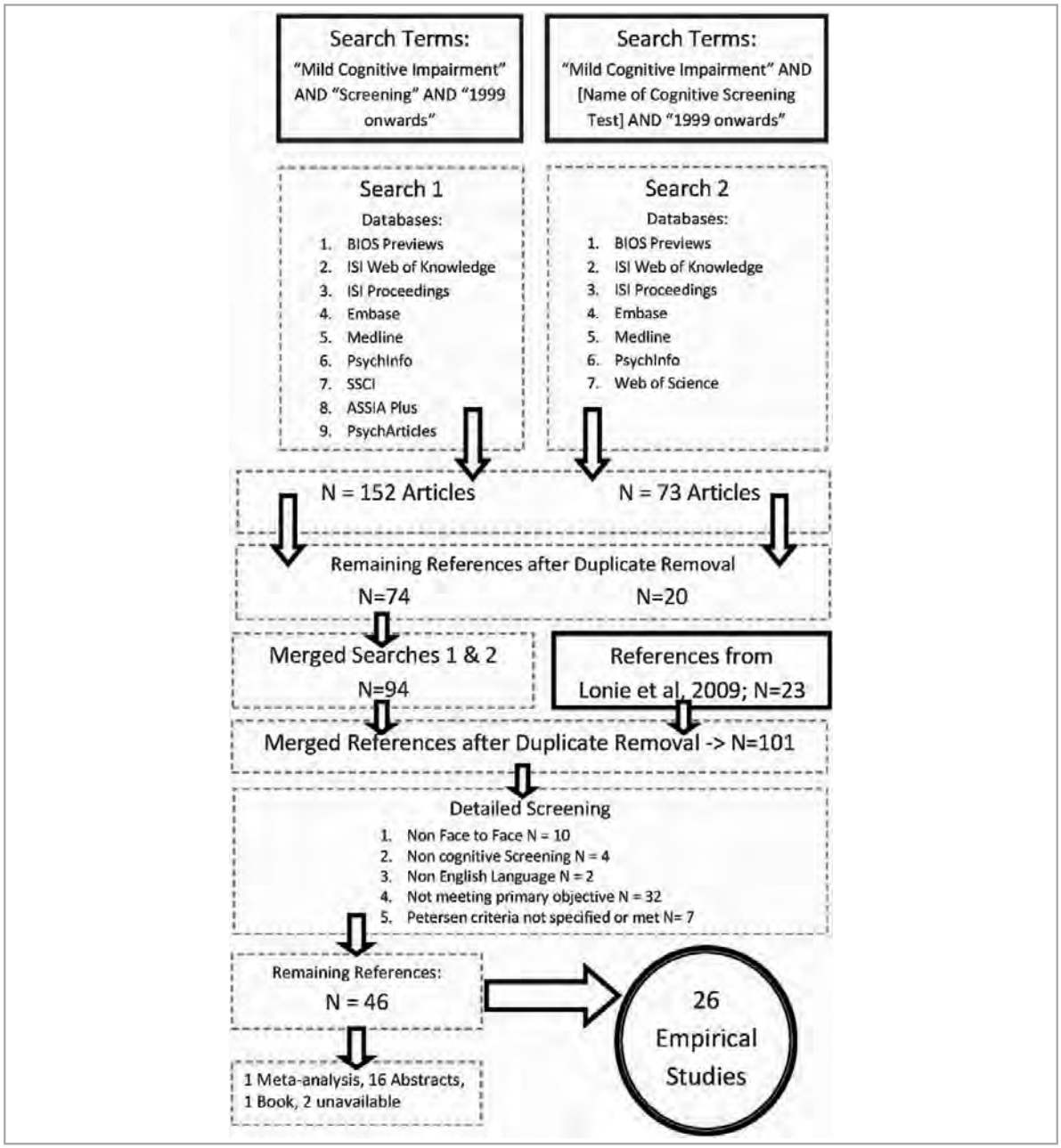


Figure 1. Search and attrition chart

been tested for screening in community samples, as their longer application time would predict.

Non-comprehensive cognitive screening

Sensitivities over 80% for detecting aMCI among healthy volunteers were reported for only three of the 11 non-comprehensive screening tests. Equivalent specificity values over 80% could be found for five non-comprehensive tests. In only six studies were aMCI sample sizes >100, and prognostic values were reported for only one of the 11identified screening measures. **Table 1** summarizes the relevant characteristics of the cognitive screening measures with incomplete domain coverage.

The MMSE [45] is the most commonly used screening test in clinical practice [46]. A recent meta-analysis of its accuracy in the detection of dementia and aMCI [3] concluded that it

Mind & Brain, the Journal of Psychiatry jop218262.3d 5/3/10 19:50:04
The Charlesworth Group, Wakefield +44(0)1924 369598 - Rev 9.0.225/W (Oct 13 2006)

Table 1. Studies included, MCI Patient Characteristics, and Diagnostic Sensitivity and Specificity compared with Healthy Volunteers

		Values for mild cognitive impairment patients							
Cognitive screening measure	**Reference**	**N**	**Mean age (SD) (years)**	**Mean years of education (SD)**	**Mean MMSE (SD)**	**Sensitivity**	**Specificity**	**Area under the curve (SE)**	**Cutoff value**
Comprehensive tests									
ACE-R	Mioshi 2006 [29]	36	68.8 (9.0)	12.8 (3.4)	27.7 (1.5)	0.84	1.00		82/100
CAMCOG	Marcos 2006 [28]	82	77.6 (6.1) (CV)	84.2% <10 (CV)	25.8 (2.8) (CV)	0.92*	0.72*		79.5
			72.9 (6.8) (NCV)	77.3% <10 (NCV)	29.4 (3.5) (NCV)				
	Diniz 2008 [15] Amnestic MCI	25	72.3 (6.6)	10.8 (5.2)	26.40 (3.0)	0.79	0.64	0.74 (0.05)	94/95
	Diniz 2008 [15] Non-amnestic MCI	11	67.8 (4.8)	9.63 (6.6)	27.7 (1.8)	0.78	0.72	0.78 (0.06)	94/95
	Diniz 2008 [15] Multidomain MCI	51	70.2 (6.5)	8.5 (4.4)	26.5 (2.4)	0.85	0.75	0.84 (0.03)	91/92
	Heinik 2009 [17]	22 (27% aMCI, 50% m-aMCI, 9% sMCI, 13.7% m-non-aMCI)	77.1 (7.5)	9.8 (4.5)	26.5 (2.1)	0.87	0.86	0.90 (0.05)	91/92
MoCA	Nasreddine 2005 [41]	94	75.19 (6.27)	12.28 (4.32)	27.00 (1.8)	0.90	0.87		26/27
	Lee 2008 [16]	37	71.3 (5,9)	8.3 (3.8)	24.0 (2.9)	0.89	0.84	0.94 (0.02)	22/23
CERAD	Karrasch 2005 [25]	15	67.50 (9.2)	8.2 (2.1)	26.50 (2.3)	N/A	N/A		N/A
	Chandler 2007 [21]	60 Community sample	72.80 (7.5)	14.80 (2.8)	27.50 (1.8)	0.81	0.72		85.1/100
Non-comprehensive tests									
CDT	Beinhoff 2005 [18]	48	66.4 (7.1)	1.87† (0.94)	28.3 (1.5)	0.40		0.57	>1
	Sager 2006 [36]	69	78.3 (6.7)	13.3 (3.0)	27.3 (1.9)	0.2	0.88		<8
	Ravaglia 2005 [35]	sMCI=18	76.3 (7.4)	6.3 (3.4)	25.4 (2.6)	0.26	0.85		6 or below
		aMCI=38	76.5 (7.1)	8.0 (4.4)	25.7 (3.1)	0.06			6 or below
		mMCI=57	78.6 (8.0)	5.7 (3.3)	24.3 (2.8)	0.40			6 or below
	Yamamoto 2004 [40]	48	74.7 (6.3)	11.5 (3.7)	27.2 (2.1)	0.75‡	0.76‡		7
	Heinik 2009 [17]	22 (27% aMCI, 50% m-aMCI, 9% sMCI, 13.7% m-non-aMCI)	77.1 (7.5)	9.8 (4.5)	26.5 (2.1)	No significant cutoff			
DemTect	Kalbe 2004 [24]	97	72.1 (9.0)	10.0 (1.2)	26.9 (1.5)	0.80	0.92		13
CAMCI	Lam 2008 [26]	S1=182	78.9 (7.0)	2.8 (4.2)	23.9 (3.2)	NR	NR	S1= 0.91‡	15/16
		S2=162	72.8 (6.5)	3.1 (3.4)	24.5 (2.7)			S2= 0.98‡	
		2 Community samples							

Mind & Brain, the Journal of Psychiatry jop218262.3d 5/3/10 19:50:05
The Charlesworth Group, Wakefield +44(0)1924 369598 - Rev 9.0.225/W (Oct 13 2006)

Table 1. Continued

		Values for mild cognitive impairment patients							
Cognitive screening measure	**Reference**	**N**	**Mean age (SD) (years)**	**Mean years of education (SD)**	**Mean MMSE (SD)**	**Sensitivity**	**Specificity**	**Area under the curve (SE)**	**Cutoff value**
ECR	Saka 2006 [37]	18	69.4 (8.3)	8.4 (5.0)	26.6 (1.7)	0.56	0.79	0.69	9 Third free recall trial
M@T	Rami 2007 [33, 34]	50	76.6 (6.6)	8.4 (5.2)	25.1 (2.4)	0.96	0.79		37
SIS	Callahan 2002 [20]	N/A	N/A	N/A	N/A	50.4	97.4		All >three errors
		Community sample							
		N/A	N/A	N/A	N/A	74.2	96.0		All >three errors
Mini-Cog	Borson 2005 [42]	71	N/A	N/A	N/A	0.58	N/A		<3
ABCS	Molloy 2005 [30]	124	77.2 N/A	12.1 N/A	27.2 N/A	0.80	0.5		N/A
	Standish 2007 [38]	166	N/A	N/A	27.1 N/A	N/A	N/A		N/A
STMS	Tang-Wai 2003 [39]	129	79.5 (7.2)	13.3 (3.2)	26.3 (2.2)	N/A‡			N/A
ADAS-cog	Pyo 2006 [43]	135	70.4 (10.0)	12.8 (3.1)	26.4 (2.6)	0.73	0.89		6 (Total score)
		Community sample							
	Grundman 2004 [23]	769	72.9 (7.3)	14.7 (3.1)	27.3 (1.9)	N/A	N/A		N/A
		Community sample							
	Fleisher 2007 [44]	539	74.9 (6.6) (P)	14.5 (3.1) (P)	N/A	N/A‡	N/A		N/A
		Community sample	71.5 (7.4) (NP)	14.97 (2.8) (NP)					
Animal fluency	Sager 2006 [36]	69	78.3 (6.7)	13.3 (3.0)	27.3 (1.9)	0.54	0.88	HC=0.76	<14
								MCI=0.78	<17

Table adapted, modified, and updated from [13].
*Compares converters with non-converting MCI.
†Education: 1=up to 10 years of education; 2=more than 10 and up to 13 years of education; 3=university degree.
‡Demented patients included in MCI group.
All samples are outpatient/memory clinic samples, unless marked as ''community samples''.
AD, Alzheimer's disease; aMCI, amnestic MCI; CI, all participants with cognitive impairment; CV, MCI converter to dementia; Dem, dementia; Dep, depression; ES, English speaking; FU, follow-up; mMCI, multiple domain MCI; N/A, not assessed; HC, healthy control participant; NCV, MCI non-converter to dementia; NP, non-progressive; P, progressive; S1, sample 1; SD, standard deviation; sMCI, single non-memory domain MCI.

Mind & Brain, the Journal of Psychiatry jop218262.3d 5/3/10 19:50:11
The Charlesworth Group, Wakefield +44(0)1924 369598 - Rev 9.0.225/W (Oct 13 2006)

''offers modest accuracy with the best value for ruling out a diagnosis of dementia in the community and primary care. *For all other uses it should be combined with or replaced by other methods*''. In particular, its ability to differentiate aMCI cases from healthy control subjects could not be established consistently [37, 47], and differences, where they do exist, range from less than one [35] to under two scale points [47]. **Table 1** shows that over half of aMCI patients score well above the commonly used MMSE cutoffs of 24 and 26 in most studies, so there is likely to be a considerable overlap between patients with aMCI and age-matched healthy control subjects. The sensitivity of the MMSE varies between 1% [36] and 96% [17], with a pooled mean of 62.7% [3]. This large variability will be related to different MMSE cutoff values, and in studies not included in Mitchell's meta-analysis [3], to different comparison groups, such as a combined group of aMCI and early dementia sufferers [26]. In turn, specificity for the MMSE in aMCI compared with healthy volunteers is often high, ranging from 58% to 100% [17, 36], although Mitchell [3] reports a pooled average of 63.3% in specialist settings. Tang-Wai *et al* [39] report no difference in baseline MMSE scores of healthy individuals who later developed aMCI and those who did not. Comparing the MMSE with other screening measures suggests that it is less effective in discriminating aMCI patients from healthy age-matched control subjects [24, 31, 34, 38, 39], from persons with progressive and non-progressive forms of cognitive impairment [33, 39], and is more sensitive to age and education effects [30].

Sensitivity and specificity were reported for the CDT [17, 18, 35, 36, 40], DemTect [24], M@T [34], and ADAS-Cog [32]. Sensitivity and specificity values have not been reported for the STMS [39] or the CAMCI [26]. For a further two, CDT [17, 18, 35, 36] and the ABCS [30], these values were too low to support their use in aMCI screening. The only sets of favorable values have been reported for the CDT [40] and the CAMCI [26], and these probably reflect the addition of patients in the early stages of dementia to the MCI patient sample. The SIS [20] appears to be of greater usefulness in detecting MCI among the elderly in specialty settings than in the community, although its items being restricted to orientation and memory questions only probably makes it inadequate for use in secondary and tertiary care.

Reliability data have been reported for the DemTect (inter-rater $r=0.99$ [24]) and the MoCA (1 month retest $r=0.92$ [31]), but for none of the other tests. In addition, practice effects over 6 and 12 months of less than one point have been reported for the DemTect [24]. Predictive validity based on longitudinal follow-up periods of 3–5.5 years are available for two of the measures, with correct classification rates of around 70% for both the STMS [39] and ADAS-Cog [22]. Administration times are on the whole shorter than 10 min, with a majority (CDT, M@T, SIS, ABCS, and STMS) reporting minimum administration times of 5 min or less.

CONCLUSION

The scarcity of validated screening instruments for aMCI is problematic. The large numbers of older adults in the community who are cognitively impaired but not demented [48, 49], and the high frequency of aMCI in secondary and tertiary specialist clinics [7, 11] emphasize the need for such screening tools.

Of the 15 screening instruments identified, four (ACE-R, DemTect, MoCA, and M@T) showed adequate (i.e., ≥79%) sensitivity and specificity for aMCI [1] compared with normal elderly control subjects. Two of these (M@T and ACE-R) do not require transformation of raw scores, and all take less than 15 min to administer. There have been no comparisons of aMCI with depressed patients, which is surprising in view of the frequency of comorbid MCI and depression [50], and the difficulties clinicians face in identifying their respective contributions to cognitive impairment [51, 52]. In general, data about cross-cultural usage, reliability, and predictive validity are rare.

In secondary and tertiary care clinics, results from the four more comprehensive cognitive screening tests described here can be analyzed in detail, and can become an integral part of the clinical evaluation and differential diagnosis. Although all these measures appear to have adequate sensitivity to aMCI and early Alzheimer's disease (AD), there is little information about their reliability in healthy elderly persons, and about the predictive validity of diagnostic cutoff scores. This makes the interpretation of score changes difficult and limits our ability to supply prognostic information. While several cognitive screening instruments thus afford the clinician the ability to detect aMCI, early AD, and in some cases non-AD dementia, they cannot currently be used to make reliable inferences about the course and eventual outcome of MCI.

Future research should focus on establishing a wider range of psychometric test properties (i.e., reliability and predictive validity) for those cognitive screening measures with adequate sensitivity and specificity. This would help us to interpret score changes and provide information for prognosis. It would also be of interest to compare the psychometric characteristics of combining brief cognitive screening measures or neuropsychological measures with the screening scales described above. It is possible that some of the brief neuropsychological measures perform as well as, if not better than, cognitive screening instruments [27, 53]. Although the results reported above refer to cognitive screening instruments in the assessment of the single and multidomain amnestic MCI subtypes, it will be important to establish their usefulness in MCI's non-amnestic forms [15].

ABBREVIATIONS

ABCS	AB Cognitive Screen
ACE-R	Addenbrooke's Cognitive Examination-Revised
ADAS-cog	Alzheimer's Disease Cooperative Study—Cognitive subscale
CAMCOG	Cambridge Cognitive Examination
CDT	Clock Drawing Test
CERAD	Consortium to Establish a Registry for Alzheimer's disease
LST	Letter Sorting Test

Mind & Brain, the Journal of Psychiatry jop218262.3d 5/3/10 19:50:12
The Charlesworth Group, Wakefield +44(0)1924 369598 - Rev 9.0.225/W (Oct 13 2006)

M@T	Memory Alteration Test
MIS	Memory Impairment Screen
MMSE	Mini-Mental Cognitive Examination
MoCA	Montreal Cognitive Assessment
SIS	Six Item Screen
STMS	The Short Test of Mental Status
VF	verbal fluency

Disclosure: The authors report no biomedical financial interests or potential conflicts of interest.

Acknowledgments: This study was funded by the Gordon Edward Small's Charitable Trust, Edinburgh (Scottish Charity Register: SC008962).

REFERENCES

1. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol*. 1999; 56(3):303–308.
2. Cullen B, O'Neill B, Evans JJ, Coen RF, Lawlor BA. A review of screening tests for cognitive impairment. *J Neurol Neurosurg Psychiatry*. 2007;78:790–799.
3. Mitchell AJ. A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. *J Psychiatr Res*. 2009;43(4):411–431.
4. Busse A, Angermeyer MC, Riedel-Heller SG. Progression of mild cognitive impairment to dementia: a challenge to current thinking. *Br J Psychiatry*. 2006;189:399–404.
5. Busse A, Bischkipf J, Riedel-Heller SG, Angermeyer MC. Subclassifications for mild cognitive impairment: prevalence and predicitive validity. *Psychol Med*. 2003;33:1029–1038.
4 6. Lehrner J, Gufler R, Guttmann G, *et al*. Annual conversion to Alzheimer disease among patients with memory complaints attending an outpatient memory clinic: the influence of amnestic mild cognitive impairment and the predictive value of neuropsychological testing. *Middle Eur J Med*. 2005(117):629-635.
7. Alladi S, Arnold R, Mitchell J, Nestor PJ, Hodges JR. Mild cognitive impairment: applicability of research criteria in a memory clinic and characterization of cognitive profile. *Psychol Med*. 2006;36(4):507–513.
8. Storandt M, Grant EA, Miller JP, Morris JC. Longitudinal course and neuropathologic outcomes in original vs revised MCI and in pre-MCI. *Neurology*. 2006;67:467–473.
9. Tabert MH, Many JJ, Liu X, *et al*. Neuropsychological prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. *Arch Gen Psychiatry*. 2006;63:916–924.
10. Palmer K, Berger AK, Monastero R, Winblad B, Backman L, Fratiglioni L. Predictors of progression from mild cognitive impairment ot Alzheimer disease. *Neurology*. 2007;68:1596–1602.
5 11. Lonie JA, Herrmann LL, Donaghey CL, Ebmeier KP. Clinical referral patterns and cognitive profile in mild cognitive impairment. *Br J Psychiatry*. 2008; 192:59-64.
12. Petersen RC, O'Brien J. Mild cognitive impairment should be considered for DSM-V. *J Geriatr Psychiatry Neurol*. 2006;19(3):147–154.
13. Lonie JA, Tierney KM, Ebmeier KP. Screening for mild cognitive impairment: a systematic review. *Int J Geriatr Psychiatry*. 2009;24(9):902–915.
14. Winblad B, Palmer K, Kivipelto M, *et al*. Mild cognitive impairment—beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *J Intern Med*. 2004;256(3): 240–246.
15. Diniz BS, Nunes PV, Yassuda MS, *et al*. Mild cognitive impairment: cognitive screening or neuropsychological assessment? *Rev Bras Psiquiatr*. 2008;30(4):316–321.
16. Lee JY, Lee DW, Cho SJ, *et al*. Brief screening for mild cognitive impairment in elderly outpatient clinic: validation of the Korean version of the Montreal Cognitive Assessment. *J Geriatr Psychiatry Neurol*. 2008; 21(2):104–110.
17. Heinik J, Shaikewitz D. The Clock Drawing Test—Modified and Integrated Approach (CDT-MIA) as an instrument for detecting mild cognitive impairment in a specialized outpatient setting. *J Geriatr Psychiatry Neurol*. 2009;22(3):171–180.
18. Beinhoff U, Hilbert V, Bittner D, Gron G, Riepe MW. Screening for cognitive impairment: a triage for outpatient care. *Dementia Geriatr Cogn Disord*. 2005;20(5):278–285.
19. Borson S, Scanlan JM, Watanabe J, Tu SP, Lessig M. Simplifying detection of cognitive impairment: comparison of the Mini-Cog and Mini-Mental State Examination in a multiethnic sample. *J Am Geriatr Soc*. 2005;53(5):871–874.
20. Callahan CM, Unverzagt FW, Hui SL, Perkins AJ, Hendrie HC. Six-item screener to identify cognitive impairment among potential subjects for clinical research. *Med Care*. 2002;40(9):771–781.
21. Chandler MJ, Lacritz LH, Hynan LS. A total score for the CERAD neuropsychological battery. *Neurology*. 2007;65:102–106.
22. Fleisher AS, Sowell BB, Taylor C, *et al*. Clinical predictors of progression to Alzheimer disease in amnestic mild cognitive impairment. *Neurology*. 2007;68(19):1588–1595.
23. Grundman M, Petersen RC, Ferris SH, *et al*. Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Arch Neurol*. 2004;61(1):59–66.
24. Kalbe E, Kessler J, Calabrese P, *et al*. DemTect: a new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia. *Int J Geriatr Psychiatry*. 2004;19(2):136–143.
25. Karrasch M, Sinerva E, Gronholm P, Rinne J, Laine M. CERAD test performances in amnestic mild cognitive impairment and Alzheimer's disease. *Acta Neurol Scand*. 2005;111(3):172–179.
26. Lam LCW, Tam CWC, Lui VWC, *et al*. Screening of mild cognitive impairment in Chinese older adults—a multistage validation of the Chinese abbreviated mild cognitive impairment test. *Neuroepidemiology*. 2008;30(1):6–12.
27. Lonie JA, Herrmann LL, Tierney KM, *et al*. Lexical and semantic fluency discrepancy scores in aMCI and early Alzheimer's disease. *J Neuropsychol*. 2009;3(Pt 1):79–92.
28. Marcos A, Gil P, Barabash A. Neuropsychological markers of progression from mild cognitive impairment to Alzheimer's disease. *Am J Alzheimers Dis Other Dementias*. 2006;21(1):189–196.
29. Mioshi E, Dawson K, Mitchell J, Arnold R, Hodges JR. The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *Int J Geriatr Psychiatry*. 2006;21(11): 1078–1085.
30. Molloy DW, Standish TIM, Lewis DL. Screening for mild cognitive impairment: comparing the SMMSE and the ABCS. *Can J Psychiatry*. 2005; 50:52–58.
31. Nasreddine ZS, Phillips NA, Bedirian V, *et al*. The montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;53(4):695–699.
32. Pyo G, Elble RJ, Ala T, Markwell SJ. The characteristics of patients with uncertain/mild cognitive impairment on the Alzheimer disease assessment scale-cognitive subscale. *Alzheimer Dis Assoc Disord*. 2006;20(1):16–22.
33. Rami L, Gomez-Anson B, Sanchez-Valle R, *et al*. Longitudinal study of amnesic patients at high risk for Alzheimer's disease: clinical, neuropsychological and magnetic resonance spectroscopy features. *Dementia Geriatr Cogn Disord*. 2007;24(5):402–410.
34. Rami L, Molinuevo JL, Sanchez-Valle R, Bosch B, Villar A. Screening for amnestic mild cognitive impairment and early Alzheimer's disease with M@T (Memory Alteration Test) in the primary care population.*Int J Geriatr Psychiatry*. 2007;22(4):294–304.
35. Ravaglia G, Forti P, Maioli F, *et al*. Screening for mild cognitive impairment in elderly ambulatory patients with cognitive complaints. *Aging Clin Exp Res*. 2005;17(5):374–379.
36. Sager MA, Hermann BP, La Rue A. Screening for dementia in community-based memory clinics. *WMJ*. 2006;105:25–29.
37. Saka E, Mihci E, Topcuoglu MA, Balkan S. Enhanced cued recall has a high utility as a screening test in the diagnosis of Alzheimer's disease and mild cognitive impairment in Turkish people. *Arch Clin Neuropsychol*. 2006; 21(7):745–751.
38. Standish TIM, Molloy DW, Cunje A, Lewis DL. Do the ABCS 135 short cognitive screen and its subtests discriminate between normal cognition,

Mind & Brain, the Journal of Psychiatry jop218262.3d 5/3/10 19:50:13
The Charlesworth Group, Wakefield +44(0)1924 369598 - Rev 9.0.225/W (Oct 13 2006)

mild cognitive impairment and dementia? *Int J Geriatr Psychiatry*. 2007; 22(3):189–194.
39. Tang-Wai DF, Knopman DS, Geda YE, *et al*. Comparison of the short test of mental status and the mini-mental state examination in mild cognitive impairment. *Arch Neurol*. 2003;60(12):1777–1781.
40. Yamamoto S, Mogi N, Umegaki H, *et al*. The clock drawing test as a valid screening method for mild cognitive impairment. *Dementia Geriatr Cogn Disord*. 2004;18(2):172–179.
41. Nasreddine ZS, Phillips NA, Bedirian V. The Montreal Cognitive Assessment, MOCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;54(4):695–699.
42. Borson S, Scanlan JM, Watanabe J, Tu S, Lessig M. Simplifying detection of cognitive impairment: comparison of the mini-cog and mini-mental state examination in a multiethnic sample. *J Am Geriatr Soc*. 2005;53:871–874.
43. Pyo G, Elble RJ, Ala T, Markweel SJ. The characteristics of patients with uncertain/mild cognitive impairment on the alzheimer's disease assessment scale-cognitive scubscale. *Alzheimer's Dis Assoc Disord*. 2006;20(1):16–22.
44. Fleisher AS, Sowell BB, Taylor C, Gamst AC, Petersen RC, Thal LJ. Clinical predictors of progression to Alzheimer disease in amnestic mild cognitive impairment. *Neurology*. 2007;68:1588–1595.
45. Folstein M, McHugh P. ''Mini-mental state''. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12:189–198.
46. Shulman KI, Herrmann N, Brodaty H. Survey of brief cognitive sceening instruments. *Int Psychogeriatr*. 2006;18:281–294.
47. Slavin MJ, Sandstrom CK, Tran TTT, Doraiswamy PM, Petrella JR. Hippocampal volume and the Mini-Mental State Examination in the diagnosis of amnestic mild cognitive impairment. *Am J Roentgenol*. 2007; 188(5):1404–1410.
48. Busse A, Bischkopf J, Riedel-Heller SG, Angermeyer MC. Mild cognitive impairment: prevalence and incidence according to different diagnostic criteria: results of the Leipzig Longitudinal Study of the Aged (LEILA75+). *Br J Psychiatry*. 2003;182(5):449–454.
49. Lopez OL, Jagust WJ, DeKosky ST, *et al*. Prevalence and classification of mild cognitive impairment in the Cardio-vascular Health Study Cognition Study. *Arch Neurol*. 2003;60:1385–1389.
50. Solfrizzi V, D'Introno A, Colacicco AM, *et al*. Incident occurence of depressive symptoms among patients with mild cognitive impairment - the Italian longitudinal study on aging. *Dementia Geriatr Cogn Disord*. 2007;24:55–64.
51. Lezak MD. *Neuropsychological Assessment*. London: Oxford University Press; 1995.
52. O'Carroll RE, Curran SM, Ross M, *et al*. The differentiation of major depression from dementia of the Alzheimer type using within-subject neuropsychological discrepancy analysis. *Br J Clin Psychol*. 1994;33(Pt 1):23–32.
53. Lonie JA, Tierney KM, Herrmann LL, *et al*. Dual task performance in early Alzheimer's disease, amnestic mild cognitive impairment and depression. *Psychol Med*. 2009;39(1):23–31.

Mind & Brain, the Journal of Psychiatry jop218262.3d 5/3/10 19:50:13
The Charlesworth Group, Wakefield +44(0)1924 369598 - Rev 9.0.225/W (Oct 13 2006)

Predicting outcome in mild cognitive impairment– a 4-year follow-up study

Jane A. Lonie, Mario A. Parra-Rodriguez, Kevin M. Tierney, Lucie L. Herrmann, Claire Donaghey, Ronan E. O'Carroll, Klaus P. Ebmeier

Number of words in paper: 3290
Number of words in abstract: 234
Number of characters in title: 73
Number of References: 24
Number of tables: 3
Number of figures: 1

Search Terms: Cognitive impairment; dementia; neuropsychology; cohort study; Hopkins Verbal Learning Test

Jane A. Lonie, PhD, Division of Psychiatry, University of Edinburgh; Mario A. Parra-Rodriguez, MD, Division of Psychiatry, University of Edinburgh; Kevin M. Tierney, MSc, Division of Psychiatry, University of Edinburgh; Lucie L. Herrmann, MA, Division of Psychiatry, University of Edinburgh & Department of Psychiatry, University of Oxford; Claire Donaghey, MSc, Division of Psychiatry, University of Edinburgh; Ronan E. O'Carroll, PhD, Department of Psychology, University of Stirling; Klaus P Ebmeier, MD, Division of Psychiatry, University of Edinburgh & Department of Psychiatry, University of Oxford.
Correspondence: Professor Klaus P Ebmeier, Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford, OX3 7JX, UK, E-mail: klaus.ebmeier@psych.ox.ac.uk

Abstract

Background: Cognitive impairment precedes the diagnosis of Alzheimer's disease (AD). It is unclear which psychometric measures predict dementia, and what cut-off points should be used. Replicable cognitive measures to provide information about differential diagnosis *and* prognosis would be clinically useful.

Aims: In a prospective cohort study we investigated which measures distinguish between patients with amnestic mild cognitive impairment (aMCI) who convert to dementia and those who do not, and which combination of measures best predicts the fate of aMCI patients.

Methods: Forty-four patients with aMCI underwent extensive neuropsychological assessment at baseline and annually thereafter for an average of 4 years. Differences in baseline cognitive performance of converters and non-converters to clinically diagnosed dementia were analysed. Classification accuracy was estimated by sensitivity, specificity, positive and negative predictive values and using logistic regression.

Results; Forty one percent of participants progressed to dementia at the end of study, with a mean annual conversion rate of 11%. Most (63%) showed persisting or progressive cognitive impairment, irrespective of diagnosis. The Addenbrooke's Cognitive Examination together with the Discrimination Index of the Hopkins Verbal Learning Test –Revised (but none of the demographic indices) differentiated converters from non-converters at baseline with 74% accuracy.

Conclusions: Targeted neuropsychological assessment, beyond simple cognitive screening, could be used in clinical practice to provide aMCI patients with prognostic information and aid selective early initiation of monitoring and treatment among patients who progress towards a clinically diagnosable dementia.

Declaration of interest: Jane A. Lonie, Mario A. Parra-Rodriguez, Kevin M. Tierney, Lucie L. Herrmann, Claire Donaghey, Ronan E. O'Carroll, and Klaus P Ebmeier report no interests that a reasonable reader would want to know about in relation to the submitted work. This pertains to all the authors of the study, their spouses or partners and their children (under 18).

Introduction

Criteria for amnestic mild cognitive impairment (aMCI) have been devised in an attempt to capture the pre-clinical phase of Alzheimer's dementia (AD). There is evidence to suggest, however, that in its current form, aMCI comprises a heterogeneous group of patients,[1, 2] some who will progress to dementia with time, and others who will not.[3] To maximise the diagnostic value of aMCI, criteria should identify a homogenous group of persons with pre-clinical dementia. To this end, cognitive criteria should be defined in a manner that reflects our current knowledge of their predictive value. As the prodromal phase of AD is likely to extent beyond a two year period,[4] studies with shorter follow-up periods are liable to underestimate the risk of conversion. In fact, only two of fourteen clinic based longitudinal aMCI studies report follow-up periods beyond 3 years (Table 1).[5, 6] Moreover, low baseline levels of general cognitive functioning among aMCI participants lead to a greater chance of neuropsychological tasks predicting dementia, because of the more advanced stage of disease in the aMCI cohort. In eleven of fourteen clinic-based longitudinal studies average baseline Mini Mental State Examination (MMSE) scores for the MCI converter group fell below the higher screening cut off for dementia (i.e. 27/30). In these studies, an underlying dementia might well have been suspected on the basis of such rudimentary cognitive screening instruments alone, begging the question of the 'added value' of a fuller cognitive evaluation. The comprehensiveness of the neuropsychological battery employed and the appropriateness (on both theoretical and empirical grounds) of test selection might also be expected to influence predictive power. For example, the combination of the Paired Associate Learning test (PAL), age and the Graded Naming Test (GNT) give an overall classification accuracy of 100% over a 2.5 year follow-up interval.[7] The same classification accuracy (i.e. 100%) has been reported for the use of the PAL in combination with the Addenbrooke's Cognitive Examination (ACE)[8] and the Graded Faces Test (GFT)[9] over shorter (1 year) intervals. However, these findings have not been replicated outside the test authors' group, in larger numbers of aMCI patients, across follow-up periods extending beyond 2.5 years, and where the mean general level of cognitive functioning at baseline (as indicated by performance on cognitive screening) falls above cut off points for dementia. If replicable, such measures could be used in the neuropsychological assessment of aMCI within specialist memory clinic settings to provide information about differential diagnosis and prognosis.

Measurement of cognitive function represents just one, albeit an important, approach to detecting and diagnosing AD at a very early and pre-clinical stage. Other work has looked at the ability of imaging (for the most part MRI scanning), biomarkers (i.e. total tau, AB42 & Phospho-tau) and changes of a behavioural nature to predict the future onset of clinically diagnosable AD. A recent meta-analysis of imaging and biomarkers for AD [10] indicated some promise for the CSF markers in so far as their

overall predictive accuracy levels were similar to that of memory impairment four years prior to the point of diagnosis. Furthermore, the effect sizes for the CSF markers were largest when assessed longer before the point of diagnosis. However, atrophy of the hippocampus or other MTL structures was found to be a less accurate predictor of future AD than memory impairment, and the largest effect sizes, which are themselves likely to represent an underestimation owing to the removal of variability inherent in the inclusion of memory impairment as a selection criterion for a majority of studies, were seen in association with measures of delayed memory recall.

In this study we present a detailed neuropsychological and clinic-based cohort study, with an average of 4 years follow-up from baseline neuropsychological assessment until final review. It is the only study to date of high-functioning aMCI converters (i.e. MMSE > 27/30) extending beyond 3 years follow-up. Furthermore, it represents the first clinic based study to investigate the robustness of the GNT, GFT, and a combination of the ACE and PAL as predictors of conversion to dementia outside the original authors' research group,[8] and to report the detailed fate of aMCI non-converters in terms of their course of cognitive impairment.

Methods

Recruitment

Participants were recruited from the Edinburgh Older Adult Neuropsychology Service, which takes all tertiary referrals over 60 years from geriatricians and old age psychiatrists in the Lothian Region of Scotland. As there is no substantial private sector, these NHS referrals are likely to be representative for patients with memory complaints attending their doctor. During the study period from September 2004 to September 2007, 46 participants with aMCI could be recruited. In particular, 71 were referred from Old Age Psychiatry, 16 from Geriatric Medicine. Of these 87 patients, 41 did not respond to the invitation to attend or refused to participate in the study, leaving 46 to be investigated as follows. Further details regarding the demographic characteristics of these patients may be found in a previous publication. [11]

Procedure

Patients who fulfilled criteria for aMCI[12] (objective cognitive impairment was defined by a performance of 1SD or more below age means on two or more measures assessing a single cognitive domain) undertook an extensive battery of neuropsychological measures at baseline and were followed-up annually, regardless of whether or not they received a clinical diagnosis of dementia during the course of the study, over an average 4 year period. A 1SD cut off point on two or more episodic memory measures was used, in place of the more commonly applied 1.5SD on one or more

measures, in an attempt to minimise the likelihood of including aMCI participants with an unstable aMCI diagnosis as well as to maximise sensitivity to memory deficits within our sample with IQs higher than average aMCI participants. At the end of the study period aMCI participants were grouped in accordance with whether or not they had received a clinical diagnosis of dementia (as documented in their medical file) at any point subsequent to their initial study assessment.

Twenty four age and IQ matched healthy elderly participants also completed the full battery of 18 neuropsychological tasks providing a normative comparison group. Sixteen of these twenty four healthy participants repeated the battery in full, an average of 28 months later. The re-test data were used to established cut-off values and criteria for further classifying the neuropsychological performance of the aMCI non-converters as 'stable aMCI', 'progressive aMCI', or 'normal' at the study endpoint.

Materials, Participants & Outcome Criteria

Participant characteristics (inclusion/exclusion criteria) and neuropsychological measures have been detailed previously.[11, 13, 14] In brief, the neuropsychological battery comprised measures of pre-morbid intelligence (National Adult Reading Test; NART) episodic memory (PAL, Hopkins Verbal Learning Test –Revised; HVLT-R, Rey Complex Figure Test; RCFT), semantic memory (GFT, GNT, Category fluency; animals), visuospatial function (RCFT copy), psychomotor processing speed (Trail Making Test Part A; TMTA) and attention/executive function (Dual Task, Controlled Oral Word Association Test; COWAT F,A,S; Trail Making Test Part B; TMTB). AMCI was defined in accordance with the revised criteria set out by Petersen.[12] Demographic characteristics of the aMCI converter, aMCI non-converter and normative sample, together with their respective baseline mean performances on cognitive screening and selected neuropsychological measures are summarised in Table 2.

Statistical Analysis

Independent t-tests were conducted to compare the baseline performances of aMCI converters and aMCI non-converters on the demographic indices of age, NART FSIQ, and years of follow-up, and the neuropsychological measures: ACE total score, PAL, HVLT-R delayed recall and discrimination index (DI; a measure of accuracy of delayed recognition), GFT, category fluency and TMTB. The alpha level was adjusted to control for multiple comparisons using Holm's sequential Bonferroni correction method.[15]

Baseline ages, time of follow-up and premorbid IQ were selected as potential confounders for their established influence on risk of developing late onset dementia.[16-18] Seven neuropsychological measures (ACE, PAL, HVLT-R DI & delayed recall, GNT, GFT, and TMTB) were selected from a total of 18, on the basis of their known sensitivity to aMCI or their high predictive validity (Table 2).

Patients who were clinically diagnosed as suffering from dementia at study endpoint were identified. Patients who had not received a clinical diagnosis of dementia were further classified as "normal" or "persisting aMCI" based on their neuropsychological profile at endpoint, and "progressive" or "non-progressive" based on the longitudinal course of cognitive function during their years of study participation. "Abnormal" neuropsychological performance was defined by a performance at the 7th centile or lower *in two or more* of 18 neuropsychological tasks (this would occur by chance in 0.83~1 of 22 MCI participants without a diagnosis of dementia). MCI decline was defined by cognitive deterioration of a magnitude *seen in fewer than 2.5%* of a sample of healthy elderly over an average 28 month period on at least two measures of semantic memory or executive functioning. Cognitive domains other than episodic memory were selected for this criterion, because of the baseline floor level performances on episodic memory tasks of many aMCI participants.

Sensitivity, specificity, positive and negative predictive values, together with the overall percentage of classification accuracy in predicting conversion or non-conversion to dementia, was determined using a combination of a total score of <88/100 on the ACE or a performance of two standard deviations or more below controls on the PAL, as this combination of measures has previously been associated with 100% sensitivity and negative predictive values.[8] These values were also determined for face (GFT) and object (GNT) naming measures.

Neuropsychological measures, for which the baseline performances of aMCI converters and aMCI non-converters were significantly different, were entered simultaneously alongside the putative confounders "age", "NART FSIQ" and "years of follow-up" into a logistic regression analysis. A backward stepwise procedure using the likelihood ratio was applied to determine model content and levels of overall classification accuracy. Criteria for entry and removal were set at p=0.05 and p=0.01, respectively, using 20 iteration (SPSS 17; SPSS Inc. Chicago, Ill, USA).

Results

Forty one percent (18/44=40.9%; 95% CI = 28 – 56%) of aMCI participants received a clinical diagnosis of dementia (most often AD) at some point prior to study endpoint (i.e. on average 4.33 years after entry into the study), giving an average annual conversion rate of 11.4% (95% CI = 4 – 23%). The remaining 26/44 (59%) participants had not received a clinical diagnosis of dementia. Medical notes were missing or not accessible for the remaining 2/44 aMCI participants. For these participants, the most up to date information available at the study endpoint was that obtained at final study attendance. Of the participants who had not received a clinical diagnosis of dementia, 8/26 (31%) were stable, 10/26 (38%) progressive, and 8/26 (31%) reverted to normal, according to the criteria defined above (Figure 1).

Following adjustment of the alpha level,[15] significant differences in the baseline performances of aMCI converters and aMCI non-converters were found on the ACE ($t_{(42)} = 2.98$, $p<0.01$, $r= 0.42$) and HVLT-DI ($t_{(41)} = 2.81$, $p <0.01$, $r=0.40$).

Only one participant obtained a GNT score at baseline below the 2nd centile of our healthy elderly control group. None of the aMCI patients performed below the 2.5th centile of our control group on the GFT at baseline. Univariate sensitivity, specificity, negative (NPV) and positive (PPV) predictive values for conversion from aMCI to dementia were, therefore, based on a cut-off performance at the 7th centile (> 1.5SD below the mean) of age norms, and can be summarised as follows; GNT sensitivity: 38%, specificity: 68%, PPV: 43%, NPV: 63%; GFNT 44%, 68%, 50%, 63%. Using a combination cut off of ACE < 88/100 or PAL > 14 errors, the overall rate of classification accuracy was 68% with sensitivity: 72%, Specificity: 65%, PPV: 59%, and NPV: 77%.

Backward logistic regression with age, NART FSIQ, years of follow-up, and the neuropsychological measures for which baseline performance differentiated converters and non-converters (HVLT-DI and ACE total score), resulted in a final model, completed after five iterations, including the variables ACE total score and HVLT-DI score only, yielding an overall classification accuracy (aMCI converter vs. aMCI non-converter) of 74%, sensitivity 65%, specificity 80%, NPV 77%, and PPV 69%.

Discussion

Forty one percent of patients who met criteria for aMCI at study entry received a clinical diagnosis of dementia within the following 4 years, giving an annual conversion rate of 10% per annum, which is almost identical to the mean (9.7%) annual conversion rate obtained on averaging the findings from existing clinic based aMCI longitudinal studies of a similar (2.5 – 3.5 year) length.

Baseline performance on the ACE and HVLT-R DI were able to discriminate between future aMCI converters and non-converters at a time when general levels of cognitive functioning fell above the higher level screening cut off for dementia (i.e. >27/30 on the MMSE and > 88/100 on the ACE), and classified MCI patients in accordance with their prognostic fate with a moderate degree (74%) of overall accuracy. Differences in the baseline performances of the aMCI converter and non-converter groups on these measures could not be explained by differences in age, FSIQ or length of follow-up, as aMCI converter and non-converter groups were roughly similar, and effects persisted after controlling for each of these variables.

The average score of the converter group on the HVLT-R DI at baseline was equal to performance at the 4th centile of published age and education matched control values,[19] and the 7th centile of our own healthy elderly age and IQ matched control sample. The corresponding values for the non-converter group were the 28th and the 36th centile, respectively, implying that there is a greater risk of

conversion to dementia among a subset of aMCI patients, who are readily identifiable on the basis of published norms.

For the ACE, average scores of the converter group were equal to the 4th, those of non-converters equal to the 31st centile of published normative values,[20] and the 1st and 16th centile of our matched study control data, providing further support for the designation of 88/100 as a higher cut-off point for dementia. We suggest that use of this score is appropriate to screen for aMCI, despite the younger age group of the original ACE normative sample.

In clinical practice, the combined performances of aMCI patients on the ACE and HVLT-R DI could be used to inform decisions about the frequency of future contact / monitoring required, or in combination with additional clinical information (i.e. levels of carer rated depressive symptoms,[21] APOE 4 carrier status,[12] corroborative history, neuroimaging findings, family history and qualitative aspects of clinical presentation to decide whether or not to consider pharmacological or other interventions. The relatively small proportion of aMCI patients showing resolution of their cognitive symptoms over the time of the study also has implications for the clinical management of such patients, as a number of empirically validated methods for the cognitive rehabilitation of early stage AD have been described[22] and could theoretically be used to enhance the day to day memory functioning of patients with aMCI.

Baseline scores of the HVLT-R DI and the ACE were significant independent predictors of conversion to dementia. Closer inspection of the regression analysis reveals that the HVLT-R DI score contributes to the overall classification accuracy of the ACE by increasing negative predictive value. This implies that memory impairment of a consolidation/storage nature is generally present in cases where a diagnosis of dementia (AD, vascular dementia (VD) or mixed AD/VD) follows within 4 years.

It is possible that cued recall impairment arises closer to the point at which AD can be diagnosed clinically, often after problems with (the more difficult) free recall become apparent. The possibility that cueing may facilitate episodic recall may then disappear with disease progression, giving rise to an encoding/consolidation profile of memory impairment.

This observation has implications for the recently proposed new research criteria for AD,[23] in which the requirement for objective evidence of significantly impaired episodic memory has been elaborated. The new criteria emphasize the importance of establishing an encoding and storage deficit on the grounds that reduced benefit from cueing during recall reliably identifies prodromal AD. Our findings lend support to the specification of episodic memory impairment in this manner. However, the limited range of scores attainable using the HVLT-R DI and the resultant potential for floor effects, suggests it may not be well-suited for monitoring significant decline in episodic memory function over time.

The newer version of the ACE-R[24] incorporates a delayed cued verbal recognition element. In light of the added predictive value of the HVLT-R DI demonstrated in this study, it would seem prudent to evaluate whether or not this measure retains its prognostic contribution alongside the ACE-R.

The mean total baseline score on the ACE (87/100) for future converters fell just below the higher level cut off point for dementia. For 26% of participants, baseline ACE scores fell above the higher cut off point for dementia (i.e. 88/100) suggesting that where the ACE is used as the sole means of determining the likelihood of developing dementia over the proceeding 4 years, up to one quarter of all persons with pre-clinical dementia receive false reassurance of 'normality'. The implications of using the ACE as a sole means to determine the presence or absence of clinically significant levels of cognitive impairment are even greater, as the present findings indicate that 62% of patients who fulfil criteria for aMCI obtain scores of 88/100 or above on the ACE.

We were unable to replicate the high levels of sensitivity, specificity, positive and negative predictive values that have been previously reported in association with combined PAL and ACE scores, the GNT and the GFT.[7-9] Our need to adopt a more conservative 7th centile (-1.5 SD) cut off for the naming measures may in part reflect the longer follow-up period in ours as compared with the last study[9] (13.7 months from baseline until study endpoint). Their shorter interval until diagnosis is consistent with a greater magnitude of impairment on naming tasks in their sample. The predictive value of neuropsychological measures is likely to vary as a function of the number of years prior to diagnosis, underscoring the need for careful consideration of both the length of follow-up and the levels of baseline cognitive functioning of aMCI cohorts in different studies.

There are a number of limitations to this study: First, although a mean follow-up period of over 4 years compares well with previous clinic based studies of longitudinal outcome in aMCI, it remains possible that additional aMCI participants will go on to receive a clinical diagnosis over the longer term. Furthermore, the length of follow-up varied among aMCI participants between 1-5 years. Ideally all aMCI participants would have been follow-up for the maximum 5-year interval. Second, the high average pre-morbid IQ and select nature (i.e. tertiary referral, amnestic single and multi-domain and primarily AD endpoint diagnosis) of our aMCI cohort limits generalization of the study findings beyond groups that are characterized similarly. Third, although the predominant eventual diagnosis of dementia was of Alzheimer (n=11) or mixed (n=5) Alzheimer/vascular type, a small proportion of aMCI participants (i.e. 2) were finally diagnosed with vascular dementia. The resultant inclusion of endpoint clinical diagnosis other than that of pure AD may have influenced the predictive validity of the neuropsychological measures within our battery. It could be argued, in a more practical sense, that exclusion of aMCI participants on grounds of multiple risk factors or even retrospectively does not reflect clinical reality. There is variability in the point at which clinicians arrive at a diagnosis of dementia and the specific criteria they employ, despite a common bias towards avoiding false positive diagnoses. It remains possible that for some aMCI patients clinical diagnostic status at

the study endpoint was in part reliant on the idiosyncrasies' of one or more of the 6 attending consultants. Finally, the relatively small sample size makes independent replication essential.

Funding

Funded by the Gordon Edward Small's Charitable Trust, Edinburgh (Scottish Charity Register: SC008962).

References

1 Blossom CMS, Matthews FE, McKeith IG, Bond J, Brayne C. Early cognitive change in the general population: how do different definitions work? *J Am Geriatr Soc* 2007; 55: 1534-40.
2 Chertkow H, Nasreddine Z, Joanette Y, Drolet V, Kirk J, Massoud F, et al. Mild cognitive impairment and cognitive impairment, no dementia: Part A, concept and diagnosis. *Alzheimers Dement* 2007; 3: 266-82.
3 Visser PJ, Kester A, Jolles J, Verhey F. Ten-year risk of dementia in subjects with mild cognitive impairment. *Neurology* 2006; 67: 1201-7.
4 Amieva H, Jacqmin-Gadda H, Orgogozo JM, Le Carret N, Helmer C, Letenneur L, et al. The 9 year cognitive decline before dementia of the Alzheimer type: a prospective population-based study. *Brain* 2005; 128: 1093-101.
5 Fox NC, Warrington EK, Seiffer AL, Agnew SK, Rossor MN. Presymptomatic cognitive deficits in individuals at risk of familial Alzheimer's disease. *Brain* 1998; 121: 1631-9.
6 Tabert MH, Manly JJ, Liu XH, Pelton GH, Rosenblum S, Jacobs M, et al. Neuropsychological prediction of conversion to Alzheimer disease in patients with mild cognitive impairment. *Arch Gen Psychiatry* 2006; 63: 916-24.
7 Blackwell AD, Sahakian BJ, Vesey R, Semple JM, Robbins TW, Hodges JR. Detecting Dementia: Novel Neuropsychological Markers of Preclinical Alzheimer's Disease. *Dement Geriatr Cogn* 2004; 17: 42-8.
8 Ahmed S, Mitchell J, Arnold R, Nestor PJ, Hodges JR. Predicting rapid clinical progression in amnestic Mild Cognitive Impairment. *Dement Geriatr Cogn* 2008; 25: 170-7.
9 Thompson SA, Graham KS, Patterson K, Sahakian BJ, Hodges JR. Is Knowledge of Famous People Disproportionately Impaired in Patients With Early and Questionable Alzheimer's Disease?*1. *Neuropsychology* 2002; 16: 344-58.
10 Schmand B, Huizenga HM, van Gool WA. Meta-analysis of CSF and MRI biomarkers for detecting preclinical Alzheimer's disease. *Psychol Med* 2010; 40: 135-45.
11 Lonie JA, Herrmann LL, Donaghey CL, Ebmeier KP. Clinical referral patterns and cognitive profile in mild cognitive impairment. *Br J Psychiatry* 2008; 192: 59-64.
12 Petersen RC, Thomas RG, Grundman M, Bennett D, Doody R, Ferris S, et al. Vitamin E and donepezil for the treatment of mild cognitive impairment. *N Engl J Med* 2005; 352: 2379-88.
13 Lonie JA, Herrmann LL, Tierney KM, Donaghey C, O'Carroll R, Lee A, et al. Lexical and semantic fluency discrepancy scores in aMCI and early Alzheimer's disease. *J Neuropsychol* 2009; 3: 79-92.
14 Lonie JA, Tierney KM, Herrmann LL, Donaghey C, O'Carroll RE, Lee A, et al. Dual task performance in early Alzheimer's disease, amnestic mild cognitive impairment and depression. *Psychol Med* 2009; 39: 23-31.
15 Holm S. A simple sequential rejective multiple test proceedure. *Scand J Stat* 1979; 6: 65-70.
16 Mebane-Sims I. 2009 Alzheimer's disease facts and figures. *Alzheimers Dement* 2009; 5: 234-70.
17 Whalley LJ, Starr JM, Athawes R, Hunter D, Pattie A, Deary IJ. Childhood mental ability and dementia. *Neurology* 2000; 55: 1455-9.

18 Petersen RC, Smith GE, Waring SC, Ivnik RJ, Kokmen E. Mild Cognitive Impairment: Clinical Characterization and Outcome. *Arch Neurol* 1999; 56: 303-8.
19 Benedict RHB, Schretlen D, Groninger L, Brandt J. Hopkins Verbal Learning Test - Revised: Normative data and analysis of inter-form and test-retest reliability. *Clin Neuropsychologist* 1998; 12: 43-55.
20 Mathuranath PS, Nestor PJ, Berrios GE, Rakowicz W, Hodges JR. A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. *Neurology* 2000; 55: 1613-20.
21 Lu PH, Edland SD, Teng E, Ingus K, Petersen RC, Cummings JL. Donepezil delays progression to AD in MCI subjects with depressive symptoms. *Neurology* 2009; 72: 2115-21.
22 Clare L, Woods RT. Cognitive training and cognitive rehabilitation for people with early-stage AD: A review. *Neuropsychol Rehabil* 2004; 14: 401.
23 Dubois B, Feldman HH, Jacova C, DeKosky ST, Barberger-Gateau P, Cummings J, et al. Reserach criteria for the diagnosis of AD: revising the NINCDS-ADRDA criteria. *Lancet Neurol* 2007; 6: 734-46.
24 Mioshi E, Dawson K, Mitchell J, Arnold R, Hodges JR. The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *Int J Geriatr Psychiatry* 2006; 21: 1078-85.

Table 1: Clinical neuropsychological studies predicting dementia in MCI

Author (Date) Reference list available from the authors	*Predictors*	*FU yrs*	*Mean MMSE/ACE at BL*	*Mean Age*	*n*	*Mean NART FSIQ Mean education level (years)*	*Cut off*	*% total C (% C per annum)*	*Results*
Thompson (2002)	GFNT	1.1	27.4	66.6	28	117 / 13	-2sd	25 (22)	GFNT PPV=0.6, NPV=0.94 Sens=0.86, Spec=0.81 GNT PPV=1.00, NPV=0.78 Sens=0.14, Spec=1.00
Ahmed (2008)	ACE ALB buildings & patterns	1	C=25.7 / 77.3 NC=29 / 86	71	18	C=11.9 NC=14	<88 >14 errors i.e. -2sd	39 (39)	ACE and / or PAL Sens=1.00, Spec=0.82 PPV=0.78, NPV=1.00
Lehmer (2005)	MMSE, Block span [AKT], Digit symbol, Misplaced objects, Name face association, SR total, SR delay	2	C=25.8 NC=28	C=71 NC=66	107	C=10 NC=12	<7	40 (20)	Selective Reminding Test delayed recall Sens >0.80, Spec >0.80 AUC = 0.94, PPV < 0.40
Griffith (2006)	DRS total Semantic fluency DRS memory VR II VR % retention DRS init/perseveration	2	28	C=70 NC=67	49	13	<37 <26%	34.22 (17)	Dementia Rating Scale initiation/perseveration score & WMS-III VR % retention Classification accuracy = 0.86, Sens=0.77 , Spec=0.89
Amieva (2004)	Age, MMSE total , MMSE word recall, BVRT, IST, DSST, LCT	2	27	C=73 NC=68	90	89% primary school			LCT only standalone predictor in regression model

						diploma			
Schmidtke (2007)	None	1.6	C=25.7 NC=26.6	C=76 NC=73	75	C=10.2 NC=9.6	NA		NA
Perri (2007)	Word list recall all indices	2	C=26.3 NC=27.7	C=73 NC=68	190	C=7.5 NC=7.7	-1.5 sd	41.5 (20.8)	Cumulative delayed recall index sens 0.75, spec 0.69
Tabert (2006)	SRT, WMS-VR, BVRT recog, BNT, ANT, BD, OA, Digit symbol, CFL, Similarities, Mattis identities and oddities	3.5	C=26.1 NC=28.1	C=73 NC=64	115	C=13.9 NC=15		48 (13.7)	Selective Reminding Test total immediate recall score Digit symbol time to completion Sens 0.76, Spec 0.90 Accuracy 0.86 PPV 0.76, NPV 0.90
Lee (2006)	CERAD word list subtests and constructional recall subtest MMSE CDR	3	C=25 NC=NA	C=71 NC=74	72			19.4 (6)	NA
Loewenstein (2007)	SIT OME	3	26	77	76	C=14 NC=12	<4	35.5 (12)	SIT recall PPV=0.70 , NPV=0.74 AUC=0.78
Estevez-Gonzalez (2004)	MMSE Age Face naming	2	C=26.3 NC=27.9	C=73 NC=66	53	C=7.1 NC=8.1	NA	48 (24)	NA
Albert (2001)	CVLT total CSRT	3	29	72	123	14	NA	19	TMTB, WMS-R VR immediate recall figures,

	TMTB SOT Alpha span FAS							(6.3)	SOT total score overall accuracy= 0.80 Sens = 0.74 , Spec = 0.83
Fox (1998) Pre-clinical familiar AD	Performance IQ RMT words	6	C=29 NC=29	44	63	100	NA	16	NA
Blackwell (2004)	NART, ADAS-cog, MMSE, RMT words, RMT faces, CANTAB Pattern recognition, Doors recognition, CANTAB delayed matching to sample, CANTAB PAL, WMS-R, LMII, GNT, New semantic naming battery, Category fluency	2.5	C=25 NC=29	C=72 NC=62	43	C=117 NC=119	NA	26 (10.4)	PAL and age PPV= 0.81 , NPV=0.97 PAL, age and GNT overall accuracy=1.00
Estevez-Gonzalez (2003)	All indices of the RAVLT except learning	2	C=26.3 NC=27.9	C=73 NC=66	70	C=7.1 NC=8.1	NA		NA
Rami (2007)	aMCI +prAD (age, visual memory) aMCI only (delayed memory test, animal fluency)	1	26 aMCI 24 prAD	73	48	7-8	NA	20 (20)	Logistic regr aMCI + prAD visual memory and age were sig predictors Logistic regr aMCI alone no sig predictors of C No details of model or accuracy of prediction provided

ACE, Addenbrooke's Cognitive Examination; ADAS, Alzheimer's Disease Assessment Scale; AKT, The Alters-Konzentrations Test; ALB, Associative learning battery; ANT, Animal Naming Test; BD, Block Design; BNT, Boston Naming Test; BVRT, Benton Visual Retention Test; CANTAB, Cambridge Automated Neuropsychological testing Assessment Battery; CERAD, Consortium to Establish a Registry for Alzheimer's Disease; CFL, lexical fluency using these letters; C, converter; CSRT, Cued and Selective Reminding test; CVLT, California Verbal Learning Test; DRS, Dementia Rating Scale; DSST, Digit Symbol Substitution Test; FAS, lexical fluency using these letters; FU, follow-up; GNT, Graded Naming Test; GFNT, Graded Faces Naming Test; IST, Isaacs Set Test; LCT, Letter Cancellation Test; MMSE, Mini Mental State Examination; n, sample size; NC, non-converter; OA, Object Assembly; OME, Object Memory Evaluation; RAVLT, Rey Auditory Verbal Learning Test; RMT, Recognition Memory Test; SIT, Semantic Interference Test; SOT, Self Ordering Test; SRT, Selective Reminding Test; TMTB, Trail Making Test Part B; VR, Visual Reproduction; WMS, Wechsler Memory Scale; reg, regression; AUC, area under the curve; NPV, negative predictive value; PPV, positive predictive value; Spec, specificity; Sens, sensitivity; Conv, converter; Non-conv, non-converter; prAD, prodromal Alzheimer's Disease.

Table 2: Demographic and baseline neuropsychological data - healthy elderly controls, aMCI converters to dementia and aMCI non-converters

Variable (maximum score)	Healthy Elderly Controls Mean (SE)	aMCI Converters Mean (SE)	aMCI Non-converters Mean (SE)	t-test Converters vs. Non-converters	Degrees of freedom	Two-tailed p-value
Demographic Information (Confounders)						
Age	70.8 (7.8)	76.0 (1.6)	73.2 (5.4)	-1.53	42	0.14
NART IQ	118.5 (3.3)	116.4 (2.0)	117.4 (1.3)	0.44	41	0.66
Months of follow-up	28.0 (9.1)	51.4 (3.0)	52.4 (2.7)	0.23	42	0.82
Cognitive Screening						
ACE total (100)	94.5 (3.2)	86.6 (1.3)	91.3 (0.9)	2.98	42	0.006~
Episodic Memory						
PAL 6 box errors^	7.9 (6.7)	23.2 (3.2)	13.5 (2.5)	-2.32	42	0.02
HVLT-R delayed recall (12)	8.1 (2.7)	3.5 (0.7)	5.8 (0.7)	2.38	41	0.02
HVLT-R DI (12)	9.9 (1.8)	7.2 (0.7)	9.2 (0.4)	2.81	41	0.008 ~
Semantic Memory						
GNT (30)	23.8 (3.1)	20.3 (1.2)	21.0 (0.8)	0.56	39	0.58
GFT (30)	20.7 (3.6)	15.3 (1.3)	18.1 (0.9)	1.84	41	0.08
Attention/Executive						
TMT B^ +	88.7 (30.7)	152.7 (19.6)	100.9 (9.8)	-2.58	42	0.014

+results were replicated using nonparametric equivalent analysis i.e. Kruskal-Wallis test due to violations of the assumption of normality and homogeneity of variance; ^ higher score indicates worse performance; ~ also significant at overall $p<0.05$ after correcting for multiple comparisons;

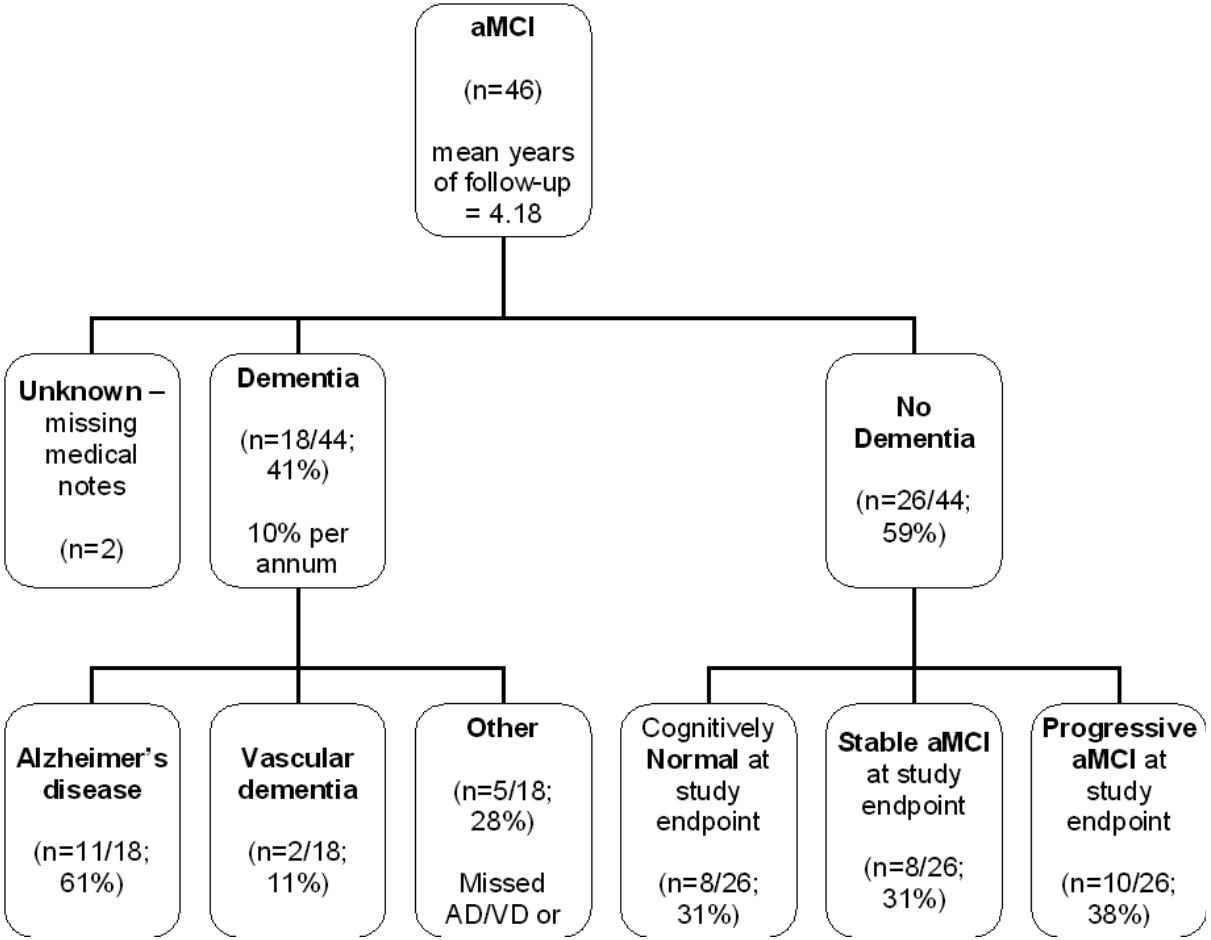
PAL = Paired Associate Learning subtest from the CANTAB battery; HVLT-R DI =Hopkins Verbal Learning Test Revised discrimination index; COWAT = controlled oral word association test; TMTB=Trail Making Test Part B; GNT=Graded Naming Test; GFT=Graded Faces Test; ACE=Addenbrooke's Cognitive Examination

Table 3: Summary of final regression model

			95% CI for exp B		
	B	SE	Lower limit	exp B	Upper limit
Constant	15.32*	6.74		4501208	
ACE Total Score	-0.15*	0.74	0.75	0.86	1.00
HVLT-R DI	-0.32*	0.16	0.53	0.73	1.00

R^2 = 0.27 (Hosmer & Lemeshow), 0.24 (Cox & Snell), 0.32 (Nagelkerke). Model χ^2 = 11.45, *p<0.05

Legend Figure 1: Flowchart of aMCI endpoint classification in accordance with clinical diagnoses in medical notes



Appendix B Tables

Table 8.1 Diagnostic and treatment details for elderly depressive participants

Patient ID No.	Clinical Diagnosis of Major Depressive Disorder Y/N (Other)	Antidepressant Y/N	Non-pharmaceutical intervention / only Y
176	Y	Y	
191	N Dysthymic	N	Y
192	Y	Y	
202	N Dysthymic	Y	Y
211	N Anxiety	N Anxiolytics	
221	Y	Y	
234	N Bipolar	Y	
235	Y	Y	
249	N Dysthymic	Y	
264	Y	N Anxiolytics	
266	Y	Y	
282	N Dysthymic	Y	
339	Y	Y	
351	Y	Y	
380	Y	Y	
381	Y	Y	
382	Y	Y	
383	Y	Y	
384	Y	N	Y
385	Y	Y	
Totals	14 Major Depressive Disorder; 4 Dysthymic; 1 Bipolar; 1 Anxiety	16 antidepressant medication; 2 anxiolytics; 2 non-pharmaceutical intervention only	3 non-pharmaceutical intervention only; 17 drug treatment

Table 8.2 Assessment battery completed as a function of participant group and stage of follow-up

	Baseline Assessment	1st 12 monthly follow-up	2nd 12 monthly follow-up	3rd 12 monthly follow-up	4th 12 monthly follow-up
MCI Assessment Protocol	n=46 NART, MMSE, HVLT-R, PAL, RCFT, TMTA&B, DUAL TASK, COWAT, Semantic Fluency, GNT, GFT, BNT, EENT	n=44 MMSE, HVLT-R, PAL, RCFT, TMTA&B, DUAL TASK, COWAT, Semantic Fluency, GNT, GFT, BNT, EENT	n=35 MMSE, HVLT-R, PAL, RCFT, TMTA&B, DUAL TASK, COWAT, Semantic Fluency, GNT, GFT, BNT, EENT	n=19 MMSE, HVLT-R, PAL, RCFT, TMTA&B, DUAL TASK, COWAT, Semantic Fluency, GNT, GFT, BNT, EENT	n=2 MMSE, HVLT-R, PAL, RCFT, TMTA&B, DUAL TASK, COWAT, Semantic Fluency, GNT, GFT, BNT, EENT ADLs PSMS Everyday functioning
Control Assessment Protocol	n=24 NART, MMSE, HVLT-R, PAL, RCFT, TMTA&B, DUAL TASK, COWAT, Semantic Fluency, GNT, GFT, BNT, EENT	n=16 MMSE, HVLT-R, PAL, RCFT, TMTA&B, DUAL TASK, COWAT, Semantic Fluency, GNT, GFT, BNT, EENT			
Early AD & Depression Assessment Protocol	n=20 NART, MMSE, HVLT-R, PAL, RCFT, TMTA&B, DUAL TASK, COWAT, Semantic Fluency, GNT, GFT, BNT, EENT				

MCI, Mild Cognitive Impairment; AD, Alzheimer's Disease; NART, National Adult Reading Test; MMSE, Mini Mental State Examination; HVLT-R, Hopkins Verbal Learning Test – Revised; PAL, Paired Associate Learning Test; RCFT, Rey Complex Figure Test; TMT, Trail Making Test; COWAT, Controlled Oral Word Association Test; GNT, Graded Naming Test; GFT, Graded Faces Naming Test; BNT, Boston Naming Test; EENT, Edinburgh Exemplar Naming Test.

Table 8.3: Summary of information from aMCI medical notes at final follow-up.

MCI Participant Number	Anticholinesterase (Y/N) (i.e. Donepezil, Rivastigmine, Reminyl, Acicept, Galantamine, memantine)	Dose of Anticholinesterase (i.e. 5mg or 10mg)	Antidepressant (Y.N)	Date of Last Contact with medical services	Frequency of service use over past 12 (months)	Type of service use (i.e. who seeing)	Medical Conditions/ co-morbid diagnoses	Dementia diagnosis Y/N (type)
1.	Cannot locate medical notes							
2.	Donepezil	5	No	18/12/08	4	Psychiatry & Neuropsychology	No	Yes
3.	No	No	No	09/05/06	0	Psychiatry & Neuropsychology	No	No
4.	No	No	No	23/02/04	0	Neuropsychology	Chronic Cardiovascular Disease	No
5.	Memantine	10	No	11/2008	1	Psychiatry	Cardiovascular disease (cause of death)	Yes (VD)
6.	No	No	No	06/2007	0	Psychiatry	No	No
7.	Notes destroyed by MRD			16/04/03	0	Neuropsychology		
8.	Rivastigmine	3 (twice day)	No	13/02/09	2	CPN	No	Yes (AD)
9.	Rivastigmine	1.5 (discontinued due to adverse effects)	No	27/03/2209	10	Psychiatry Neuropsychology Haematology	Cardiovascular Disease with history of previous myocardial infarction	Yes (Mixed VD-AD dementia)
10.	Rivastigmine	trial for 1.5 mg (28/11/03)	No	12/05/05	0	Neuropsychology	Dysphagia, Macular degeneration	Yes (AD)
11.	No	No	No	29/11/04	0	Neuropsychology	Bowel Cancer	No (suspicion of AD)
12.	Yes but not sure which one		No	14/06/07	0	Neuropsychology	Essential hypertension	No/ normal
13.	No	No	Citalopram 20 mg	10/05/05	0	CPN	Depression	No
14.	No	No	No	12/11/08	3	Psychiatry & Neuropsychology	No	No (MCI)
15.	Donepezil	10	No	24/07/2008	1	CPN	Cardiovascular Disease and High Blood Presure	Yes (AD)
16.	No	No	No	20/08/08	6	CPN, Psychiatry, Surgeon	Breast cancer and Cardiovascular disease	No
17.	No	No	No	25/11/03	0	Neuropsychology	No	No

18.	was suggested when last seen		No	17/12/08	1	Psychiatry	Cardiovascular Disease	Yes (AD)
19.	No	No	No	30/05/07	0	CPN	Pulmonary embolims, Chrohn's Disease, hip replacement	MCI
20.	No	No	No	01/11/04	0	Neuropsychology	Macular degeneration, High Blood presure, Supressed TSH, malignant neoplasm of trachea/lungs	No
21.	No	No	No	12/03/08	1	Psychiatry	No	No
22.	No	No	No	25/08/08	2	Psychiatry & Neuropsychology	hypertension and hyperlipidemia	No
23.	Yes					Nursing Home Care	NA	Yes
24.	No	No	No	29/04/03	0	Neuropsychology	No	No
25.	Memantine	10	No	08/01/08	0	Psychiatry	High Blood Pressure and Hypercholesterolemia	No (CVD)
26.	Yes	No	Citalopram 10 mg	14/05/09	0	Neuropsychology	Depression	Yes (AD or VD)
27.	No	No	No	21/04/06	0	Psychiatry	No	No
28.	No	No	Venlafaxine 75 mg ncte and 37.5 morning	14/10/04	0	Psychiatry	No	No
29.	No	No	No	13/11/06	0	Neuropsychology	No	No
30.	Galantamine	24	No	07/2008	1	CPN	Angina, hypertension, IHD, Hyperlipidemia, DM	Yes (AD)
31.	No	No	No	26/06/03	0	Neuropsychology	Macular degeneration, High Blood presure, Supressed TSH, autoimmune primary hypothyroidism	No
32.	Galantamine	24	No	26/09/08	3	Psychiatry	No	Yes (AD)
33.	Memantine	10	No	08/05/03	0	Geriatrician	Osteogenesis Imperfecta	No
34.	No	No	No	17/09/03	0	Neuropsychology	Cardiovascular disease, MI 1989, DM type II, Hypertension, Arythmia	No
35.	Galantamine	8	Paroxetine (20 mg/day)	16/3/2009	6	CPN	No	Yes (AD)

36.	No	No	No	31/10/2007	Twice weekly day hospital	Nursing staff	No	Yes (AD)
37.	Donepezil	10	Mitrazepine	23/01/2009	No documented in case notes	In-patient hospital care	Glaucoma, Sarcoidosis	Yes (AD)
38.	Rivastigmine Stopped due to GI symptoms	No	No	21/04/2008	0	Geriatrician	TIA , ischaemic heart disease, aortic stenosis, primary hyperparathyroidism	Yes (mixed AD & VD or VD)
39.	Galantamine	12	no	28/05/08	1	CPN	B12 Deficiency	Yes (AD)
40.	Galantamine	24mg	No	24/02/2009	4	Psychiatry CPN	Ischaemic heart disease Myocardial infarction Essential hypertension Abdominal aortic aneurysm Chronic kidney disease	Yes (VD or mixed VD & AD)
41.	Galantamine	8mgs	No	21/07/08	1	Neuropsychology	hypertension	Yes (VD)
42.	Donepezil	10	No	09/10/08	7	GP	High Blood Pressure and Seizures	Yes (AD)
43.	No	No	No	20/03/07	0	Psychiatry	No	No
44.	No	No	No	17/06/04	0	Neuropsychology	No	No
45.	No	No	No	13/05/08	1	Neuropsychology	No	No
46.	No	No	No	09/05/06	0	Neuropsychology	High Blood Pressure, Colon Cancer	No

CPN, community psychiatric nurse; GP, general practitioner; AD, Alzheimer's disease; VD, vascular dementia; GI, gastrointestinal; TIA, transient ischaemic attack; MI, myocardial infarction; IHD, ischaemic heart disease;

Appendix C Figures

Figure 8.1 Individual scores on the delayed recall component of the ACE as a function of group

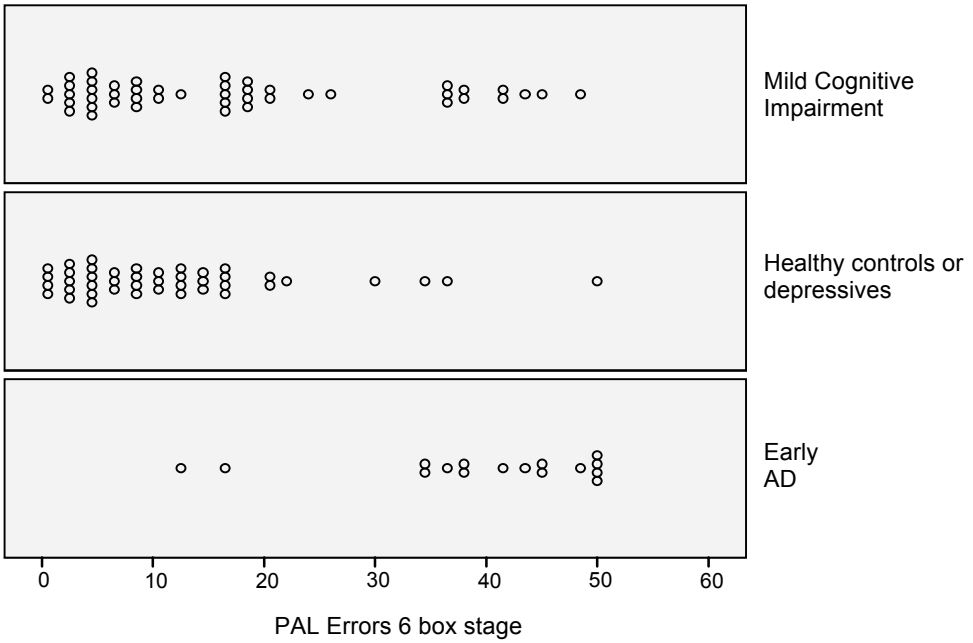


Figure 8.1 Error scores at the 6-box level of PAL as a function of group

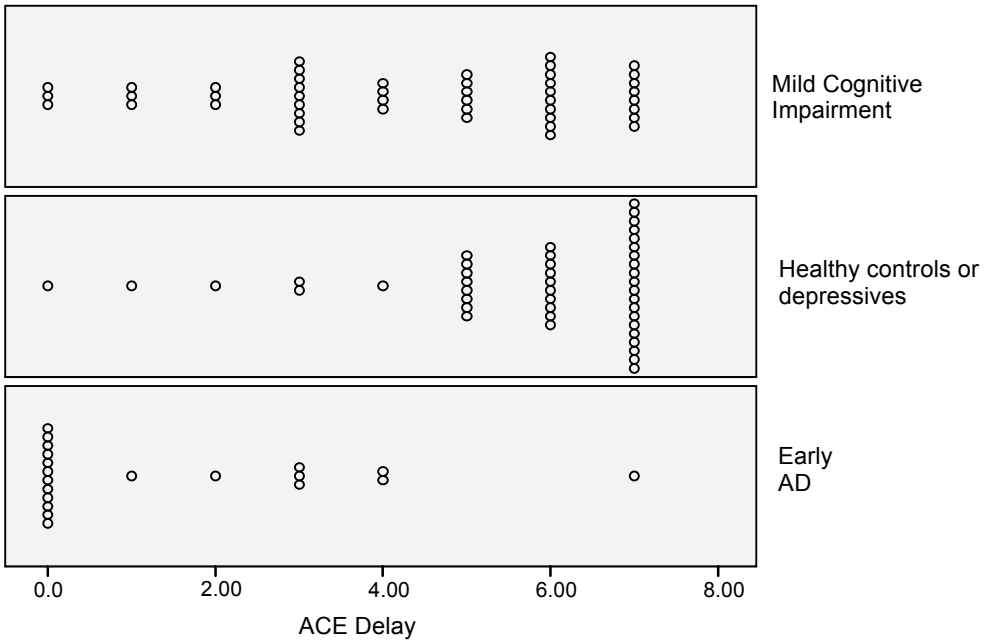


Figure 8.3 Individual total recall scores on the HVLT-R as a function of participant group

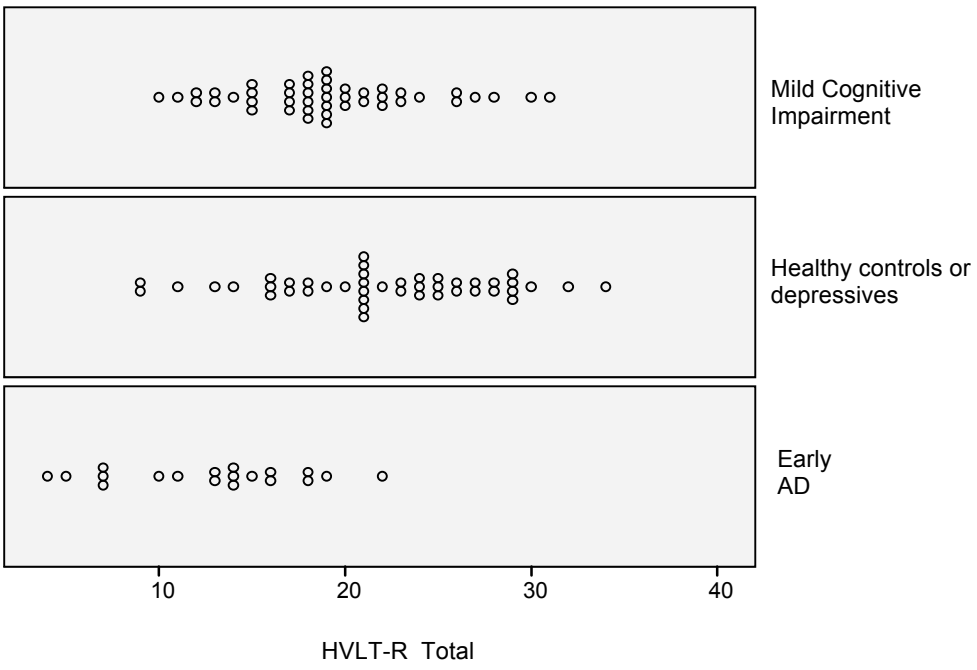


Figure 8.4 Scores on the HVLT-R delay as a function of group

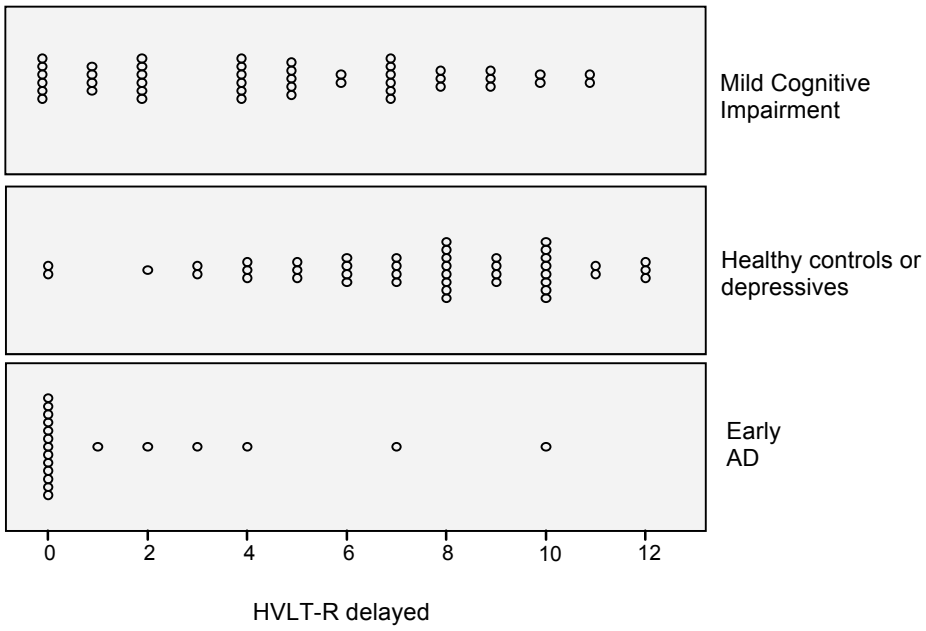


Figure 8.5 Individual scores on the HVLT-R DI as a function of group

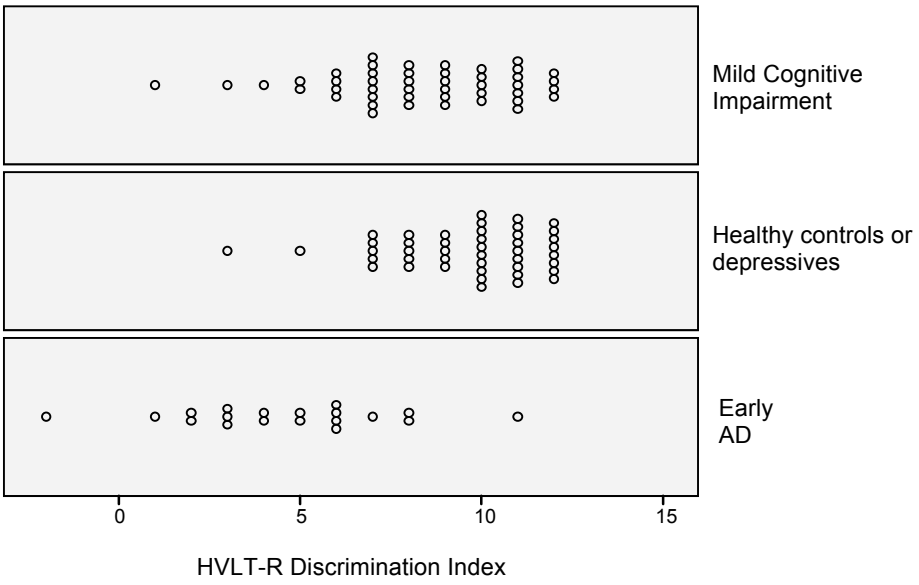
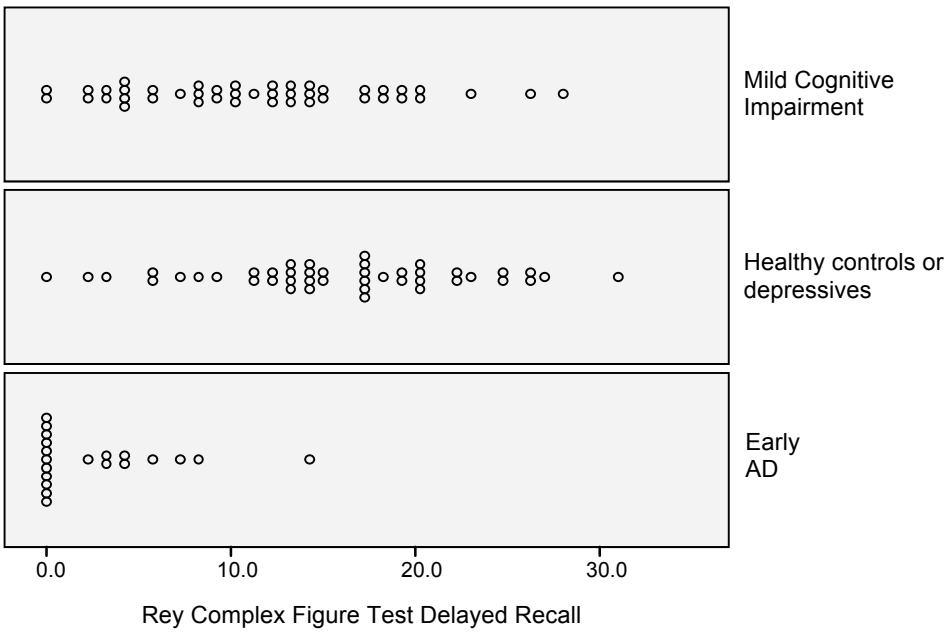


Figure 8.6 Individual scores on the delayed recall component of the Rey Complex Figure Test as a function of participant group



Appendix D Measurement Scales

(revised 2000)

ADDENBROOKE'S COGNITIVE EXAMINATION (ACE)

Name:

Date of birth:

Hospital no.:

Addressograph

Age at leaving education : _____
(school/college etc.)

Date of testing : ___ / ___ / ___

Tester's name : _____

ORIENTATION

(a) What is the Year _____
Season _____
Date ±2 _____
Record errors. Days _____
Month _____
[*Score* 0 - 5]

b) Where are we Country _____
County _____
Town _____
Record errors. Hospital/building _____
Floor *Allow if almost correct.* _____
[*Score* 0 - 5]

REGISTRATION

Name three unrelated objects, taking one second to say each: eg. **lemon, key & ball.** Say them once only and ask the patient to repeat all three. *Give one point for each correct answer at first attempt.*
If score <3 repeat the items until the patient learns all three [0 - 3]

ATTENTION / CONCENTRATION

Ask the patient to begin with **100 and and subtract 7, and keep subtracting 7.**
Stop after five subtractions (93, 86, 79, 72, 65). *Score the total number of correct subtractions.*
If score <5 : Spell **WORLD backwards.** *Score is the number of letters in the correct order, eg dlorw = 3.*
Take score of better of the two tasks. Record errors: _____ [0 - 5]

RECALL

Ask for the names of the three objects learned in question 3. *One point for each answer.* [0 - 3]

MEMORY

(a) Anterograde Memory:

Read the name and address and ask the patient to repeat it once you have finished. Regardless of the score after the first trial, repeat the task twice in exactly the same way. *Record errors at each trial.*

	1st trial	2nd	3rd	5 min delay
Peter Marshall	_____	_____	_____	_____
42 Market Street	_____	_____	_____	_____
Chelmsford	_____	_____	_____	_____
Essex	_____	_____	_____	_____
	/7	/7	/7	/7

Trial 1–3 [0 - 21] ☐
5 min delay [0 - 7] ☐

(b) Retrograde Memory:

Record errors.

Name of PM _____
Last PM _____
Opposition Leader _____
USA President _____

[0 - 4] ☐

VERBAL FLUENCY

(a) Letters — Ask the patient 'tell me as many words as you can think of beginning with the letter P, but not people and places. You have one minute to go'

(b) Animals — In the same way ask the patient to generate the names of as many animals as possible in one minute, beginning with any letter of the alphabet.

Record all responses. Error types: perseverations and intrusions.

P		Animals	
(start here)	*(continue)*	*(start here)*	*(continue)*

Animal	P	Score
>21	>17	7
17–21	14–17	6
14–16	11–13	5
11–13	8–10	4
9–10	6–7	3
7–8	4–5	2
<7	<4	1

P: Total _____ No. correct _____ [0 - 7] ☐

Animals: Total _____ No. correct _____ [0 - 7] ☐

LANGUAGE

(a) Spontaneous speech
- fluency (phrases >5 words)
- paraphasic errors (phonemic or semantic)
- word finding difficulties

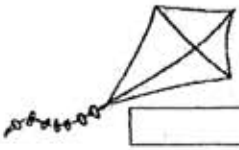
(b) Naming — Ask the patient to name the following pictures.

Record errors. [0 - 2] ☐

☐

☐

☐



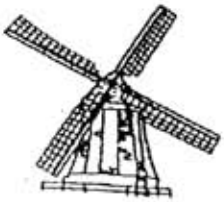
☐

☐

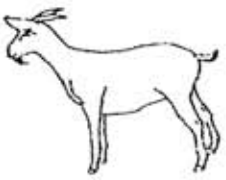
☐

☐

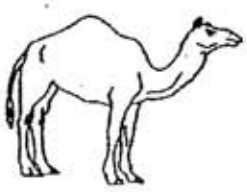
☐



☐



☐



☐

☐

[0 - 10] ☐

(c) Comprehension

single-step commands
- 'point to the door'
- 'point to the ceiling' [0 - 2] ☐
- show written instruction: [0 - 1] ☐

CLOSE YOUR EYES

3-stage command
- 'Take the paper in your hand. Fold the paper in half. Put the paper on the floor.'
  Score 1 for each correctly performed step. [0 - 3] ☐

complex grammar
- 'point to the ceiling then the door'
- 'point to the door after touching the bed/desk'
  Score 1 for each correctly performed command. [0 - 2] ☐

(d) Repetition

single words
- 'brown'
- 'conversation'
- 'articulate' [0 - 3] ☐

phrases
- 'No ifs, ands, or buts.' [0 - 1] ☐
- 'The orchestra played and the audience applauded.' [0 - 1] ☐

(e) Reading

- shed
- wipe
- board
- flame
- bridge *Score 1 if all regular words correct.* [0 - 1] ☐

- sew
- pint
- soot
- dough
- height *Score 1 if all irregular words correct.* [0 - 1] ☐

(f) Writing

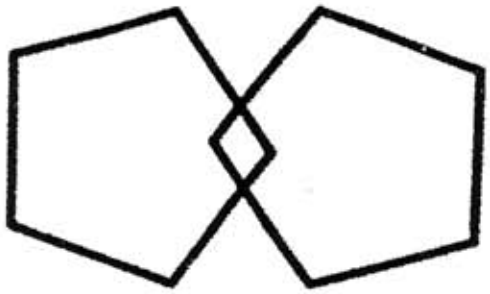
Ask the patient to make up a sentence and write it down in the space below.
If stuck, suggest a topic e.g. weather, journey to hospital.
Score 1 for a correct subject and verb in a meaningful sentence.

_____ [0 - 1] ☐

NOW CHECK delayed recall of name and address. Record errors on page 1 and enter result into box.

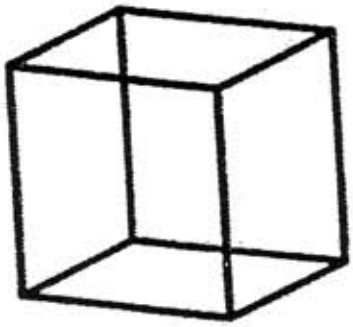
VISUOSPATIAL ABILITIES

(a) Overlapping pentagons Ask the patient to copy this diagram:



Score 1 if both figures have 5 sides and overlap. [0 - 1] [redacted]

(b) Wire cube Ask the patient to copy this drawing:



Score 1 if correct. [0 - 1] [ ]

(c) Clock Ask the patient to draw a clockface with numbers and the hands at ten past five.

Score 1 each for correct circle, numbers and hands. [0 - 3] [ ]

CHECK: Have you tested and recorded the delayed recall time of name and address (page 1)?

OVERALL SCORES MMSE: [redacted]/30 TOTAL: [ ]/100

Normative Study based on 127 subjects aged 50 to 80
Mean 93.9 ± 3.5 Cut off < 87 for age 50–80

NART Record Sheet

Participant ID: ____ *Date:* ___/___/____
Score: ___/50 *FSIQ Equialent*: _____

Chord		Superfluous	
Ache		Simile	
Depot		Banal	
Aisle		Quadruped	
Bouquet		Cellist	
Psalm		Façade	
Capon		Zealot	
Deny		Drachm	
Nausea		Aeon	
Debt		Placebo	
Courteous		Abstemious	
Rarefy		Detente	
Equivocal		Idyll	
Naïve		Puerperal	
Catacomb		Aver	
Gaoled		Gauche	
Thyme		Topiary	
Heir		Leviathan	
Radix		Beatify	
Assignate		Prelate	
Hiatus		Sidereal	
Subtle		Demesne	
Procreate		Syncope	
Gist		Labile	
Gouge		Campanile	

Column Total: ____/25 Column Total: ____/25

Dual Task

Lists for Digit Span Determination

After each of the following lists, in the space provided, enter a tick (√) if the list is correctly recalled and a cross (×) if it is not. At the bottom of the page, in the space provided, enter the subject's Digit Span as the maximum length of the lists of which the subject recalled 5/6 correctly. Present only 6 lists to the subject.

List	Result (√ or ×)	List	Result (√ or ×)	List	Result (√ or ×)
For Span = 2					
83		54		27	
28		37		91	
68		96		87	
For Span = 3					
829		687		871	
132		356		251	
152		637		915	
For Span = 4					
6241		1372		5316	
2359		7392		4815	
7132		6539		1872	
For Span = 5					
84132		85293		79514	
62143		91635		82691	
97438		16592		75468	
For Span = 6					
587261		492617		148239	
261384		247681		423896	
632147		429735		641357	
For Span = 7					
2941378		6297865		1897562	
1285394		8243167		3185624	
8693735		3945782		2473961	
For Span = 8					
65148279		28653197		85729136	
18472913		65792381		76591243	
42785921		74529638		76921358	
For Span = 9					
679174382		239874615		539748216	
746231958		867934612		513985267	
398724615		794831265		231986734	
For Span = 10					
4982176453		2853967624		2914984357	
5731298426		9781734826		6983285149	
8182397465		8491287637		6391727362	

Subject's Digit Span =

Dual task

List memory (Single Task)

Digit Span =

Note to experimenter. The table contains only lists of ten digits. The lists actually given must be equal in length to the subject's digit span. Starting from the left of each list below, read out lists of length equal to the subject's digit span. Since the lists are presented for only 1.5 minutes, the number actually read out will depend upon the subject's digit span. As the subject tries to reproduce the list, enter each item below the item that was actually in the same ordinal position when the list was read out. The raw score is the number of digits in each list that were correctly recalled in their correct serial positions. These raw scores can be converted to proportions by using the conversion table, or simply dividing by the number of lists. The subject's final List Memory score is the mean proportion, that is the total of the proportions in the rightmost column, divided by the number of lists dictated.

List	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	Score	Score/n
1.	1	5	8	7	3	6	2	9	5	4		
Response												
2	3	7	9	8	1	4	6	1	2	5		
Response												
3	6	9	3	1	4	7	5	9	8	2		
Response												
4	2	4	3	8	7	1	9	4	2	3		
Response												
5	2	1	5	3	8	6	4	7	9	6		
Response												
6	7	9	6	3	1	4	2	8	3	5		
Response												
7	8	1	6	3	9	5	7	4	2	1		
Response												
8	1	7	3	2	9	3	6	4	8	5		
Response												
9	9	6	1	2	5	3	8	2	7	4		
Response												
10	8	7	1	3	9	4	6	5	7	2		
Response												
11	3	2	1	9	5	4	3	6	8	7		
Response												
12	4	7	2	4	5	8	1	9	3	6		
Response												
13	8	4	5	1	6	2	3	4	9	7		
Response												
14	6	2	7	1	3	8	5	2	9	4		
Response												
15	8	3	9	1	6	2	7	6	5	4		
Response												

List Memory Score (Single Task) =

Dual task

List memory (Dual Task)

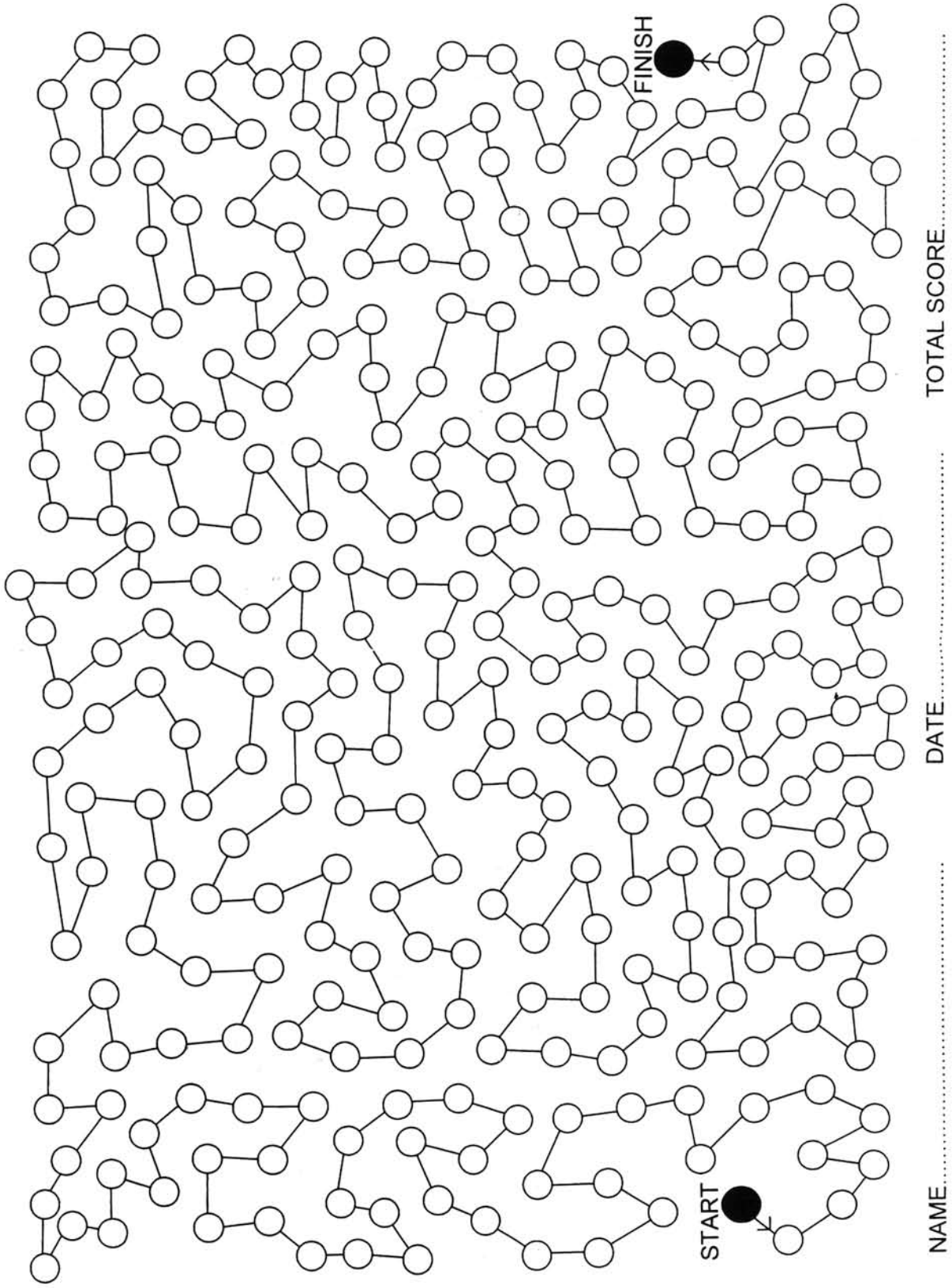
Digit Span =

Note to experimenter. The table contains only lists of ten digits. The lists actually given must be equal in length to the subject's digit span. Starting from the left of each list, read out lists of length equal to the subject's digit span. Since the lists are presented for only 1.5 minutes, the number actually read out will depend upon the subject's digit span. As the subject tries to reproduce the list, enter each item below the item that was actually in the same ordinal position when the list was read out. The raw score is the number of digits in each list that were correctly recalled in their correct serial positions. These raw scores can be converted to proportions by using the conversion table (see List Memory – Single Task), or simply dividing by the number of lists. The subject's final List Memory score is the mean proportion, that is the total of the proportions in the rightmost column, divided by the number of lists dictated.

List	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	Score	Score/n
1.	9	5	6	1	3	6	1	9	8	2		
Response												
2	7	2	9	1	5	4	8	1	6	3		
Response												
3	5	8	9	7	2	4	5	3	1	4		
Response												
4	9	6	3	8	2	5	4	7	1	8		
Response												
5	2	4	6	3	1	8	7	2	5	4		
Response												
6	5	7	8	7	2	9	4	3	5	2		
Response												
7	1	3	4	8	3	1	2	6	2	9		
Response												
8	8	2	7	5	4	6	1	3	8	9		
Response												
9	1	9	4	2	7	4	8	3	6	2		
Response												
10	3	1	2	6	9	4	8	3	5	2		
Response												
11	2	5	4	9	6	1	9	4	8	2		
Response												
12	3	8	6	4	5	7	5	2	9	6		
Response												
13	7	5	6	3	2	8	5	1	9	1		
Response												
14	9	3	5	9	6	8	2	1	3	7		
Response												
15	5	4	3	6	5	7	3	8	7	3		
Response												

List Memory Score (Dual Task) =

Dual task tracking task



BOSTON NAMING TEST

	WORD	CORRECT without cue	TIME (20')	STIMULUS CUE	TIME (20')	PHONETIC CUE
1	bed (a piece of furniture)					
2	tree (something that grows outdoors)					
3	pencil (used for writing)					
4	house (a kind of building)					
5	whistle (used for blowing)					
6	scissors (used for cutting)					
7	comb (used for fixing hair)					
8	flower (grows in a garden)					
9	saw (used by a carpenter)					
10	toothbrush (used in the mouth)					
11	helicopter (used for air travel)					
12	broom (used for cleaning)					
13	octopus (an ocean animal)					
14	mushroom (something to eat)					
15	hanger (found in a closet)					
16	wheelchair (found in a hospital)					
17	camel (an animal)					
18	mask (part of a costume)					
19	pretzel (something to eat)					
20	bench (used for sitting)					
21	racquet (used for sports)					
22	snail (an animal)					
23	volcano (a kind of mountain)					
24	seahorse (an ocean animal)					
25	dart (you throw it)					
26	canoe (used in the water)					
27	globe (a kind of map)					
28	wreath (a x-mas decoration)					
29	beaver (an animal)					

INSTRUCTIONS

- Start at item 30.
- If there is a failure before item 38, work backwards from 30 until 8 consecutive pictures are named correctly without assistance.
- Discontinue after 6 consecutive failures.
- Record responses verbatim.
- Allow 20" for a response unless there is a DK response.
- Give a stimulus cue only if a response is clearly a misperception, or lack of recognition. Allow 20" to respond after the stimulus cue has been given.
- Give a phonetic cue after every failure to respond or after any incorrect response

BOSTON NAMING TEST

WORD	CORRECT without cue	TIME (20')	STIMULUS CUE	TIME (20')	PHONETIC CUE
30 harmonica (musical instrument)					
31 rhinoceros (an animal)					
32 acorn (it comes from a tree)					
33 igloo (type of house)					
34 stilts (used to make you taller)					
35 dominoes (a game)					
36 cactus (something that grows)					
37 escalator (you go up on it)					
38 harp (a musical instrument)					
39 hammock (you lie on it)					
40 knocker (it's on a door)					
41 pelican (a bird)					
42 stethoscope (used by doctors)					
43 pyramid (found in Egypt)					
44 muzzle (used on dogs)					
45 unicorn (mythical animal)					
46 funnel (used for pouring)					
47 accordion (a musical instrument)					
48 noose (used for hanging)					
49 asparagus (something to eat)					
50 compass (for drawing)					
51 latch (part of a door)					
52 tripod (photographers use it)					
53 scroll (a document)					
54 tongs (a utensil)					
55 sphynx (it's found in Egypt)					
56 yoke (used on farm animals)					
57 trellis (used in a garden)					
58 palette (artists use it)					
59 protractor (measures angles)					
60 abacus (it's used for counting)					

INSTRUCTIONS

- Start at item 30.
- If there is a failure before item 38, work backwards from 30 until 8 consecutive pictures are named correctly without assistance.
- Discontinue after 6 consecutive failures.
- Record responses verbatim.
- Allow 20" for a response unless there is a DK response.
- Give a stimulus cue only if a response is clearly a misperception, or lack of recognition. Allow 20" to respond after the stimulus cue has been given.
- Give a phonetic cue after every failure to respond or after any incorrect response

GRADED NAMING TEST

Record Form

Pat McKenna
Elizabeth K. Warrington

NFER-NELSON

Name _____

Age _____

Patient Number _____

Date _____

No.	Object	Response	Score
1	Kangaroo		
2	Scarecrow		
3	Buoy		
4	Thimble		
5	Handcuffs		
6	Tweezers		
7	Corkscrew		
8	Sporan		
9	Tassel		
10	Sundial		

Please turn over....

No.	Object	Response	Score
11	Chopsticks		
12	Periscope		
13	Boar		
14	Blinkers		
15	Monocle		
16	Turtle		
17	Trampoline		
18	Bellows		
19	Shuttlecock		
20	Anteater		
21	Pagoda		
22	Radius		
23	Leotard		
24	Mitre		
25	Yashmak		
26	Sextant		
27	Centaur		
28	Cowl		
29	Turu		
30	Retort		
		Total Correct	

First published in 1983 by The NFER-NELSON Publishing Company Ltd., Darville House, 2 Oxford Road East, Windsor, Berks, England SL4 1DF
Printed in England by Thanet Press Limited, Margate, Kent. 2 (3.91). Code 136/014

EDINBURGH EXEMPLAR NAMING TEST

1.	ELEPHANT	_____
2.	COW	_____
3.	CROCODILE	_____
4.	EAGLE	_____
5.	FLY	_____
6.	GIRAFFE	_____
7.	FROG	_____
8.	DEER	_____
9.	GORILLA	_____
10.	BEAR	_____
11.	CAMEL	_____
12.	GOAT	_____
13.	LION	_____
14.	HIPPO	_____
15.	DRAGONFLY	_____
16.	KOALA	_____
17.	BEETLE	_____
18.	OCTOPUS	_____
19.	MONKEY	_____
20.	KANGAROO	_____
21.	JELLYFISH	_____
22.	SKUNK	_____
23.	LEOPARD	_____
24.	RHINO	_____
25.	SCORPIO	_____
26.	OWL	_____
27.	LILY	_____
28.	SLUG	_____
29.	LIZARD	_____
30.	BUTTERFLY	_____
31.	SQUIRREL	_____
32.	WHALRUS	_____
33.	TORTOISE	_____
34.	PEACOCK	_____
35.	SWAN	_____
36.	BEE	_____
37.	PANDA	_____
38.	ZEBRA	_____
39.	ROSE	_____
40.	SNAIL	_____
41.	CRAB	_____
42.	SUNFLOWER	_____
43.	SEAL	_____
44.	PENGUIN	_____
45.	TULIP	_____
46.	POLAR BEAR	_____
47.	RABBIT	_____
48.	TURKEY	_____
49.	PARROT	_____
50.	DAFFODIL	_____

HOPKINS VERBAL LEARNING TEST

Free recall

	Trial 1	Trail 2	Trail 3
LION			
EMERALD			
HORSE			
TENT			
SAPPHIRE			
HOTEL			
CAVE			
OPAL			
TIGER			
PEARL			
COW			
HUT			

Recognition

HORSE	EMERALD	balloon	apartment
house	mountain	boat	COW
HUT	CAVE	dog	LION
TENT	TIGER	HOTEL	PEARL
ruby	SAPPHIRE	coffee	penny
OPAL	cat	sea	diamond

Fluency Task

Letter Fluency

F	A	S

FAS Fluency Total = ____ P Word Fluency = ____

Category Fluency

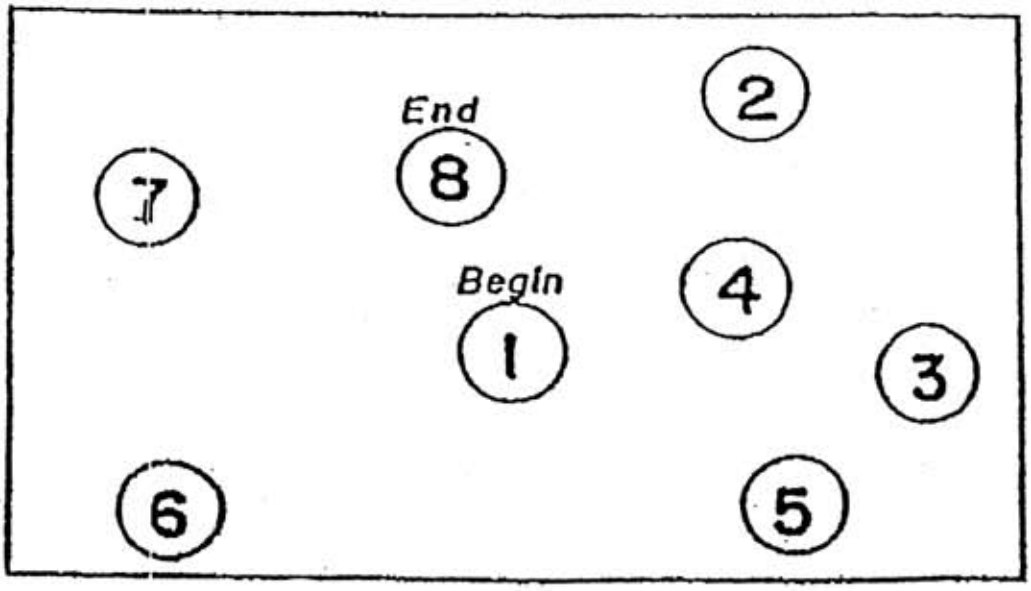
Fruit	Vegtables

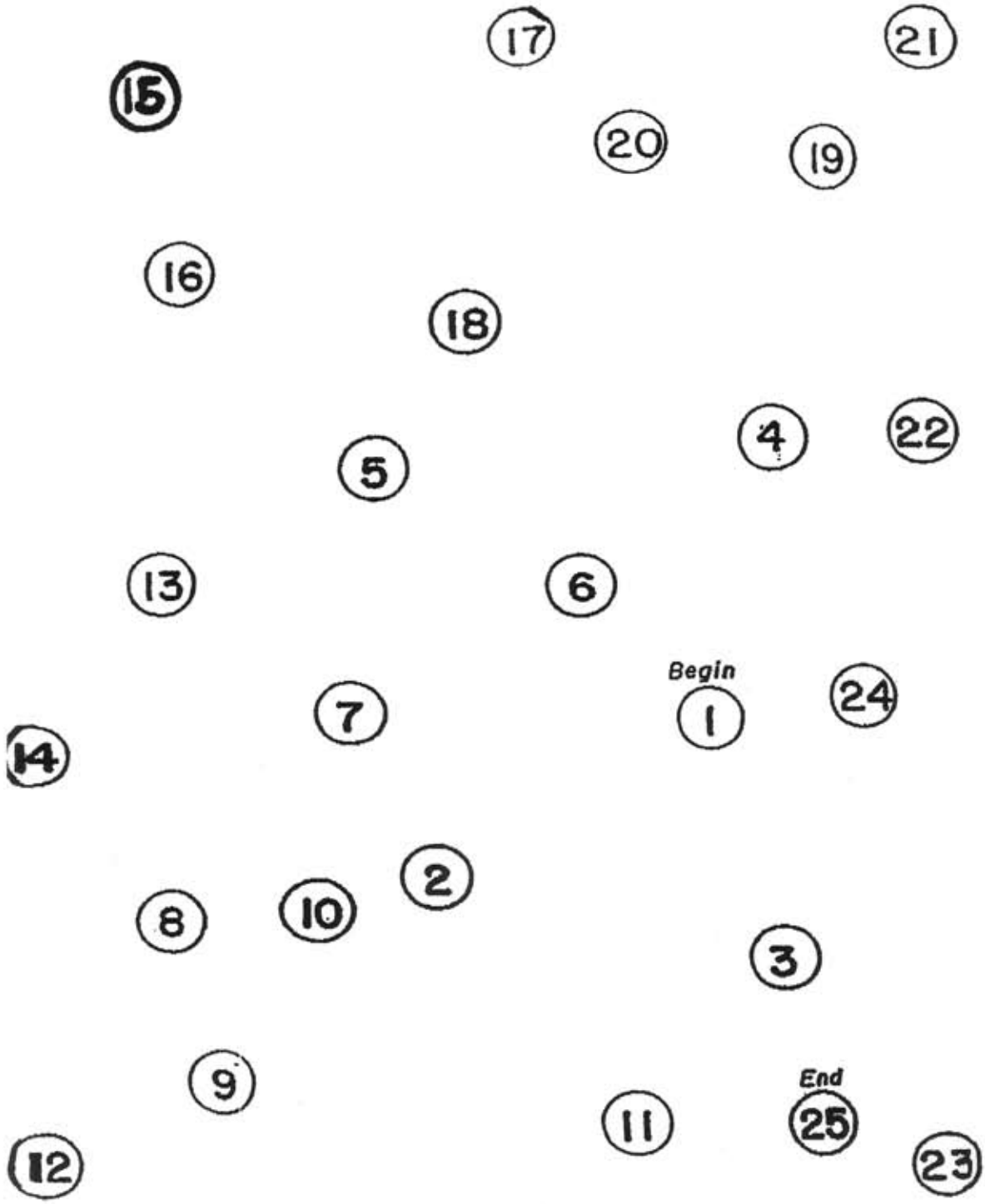
Category Fluency Total = ____ Animal Word Fluency = ____

TRAIL MAKING

Part A

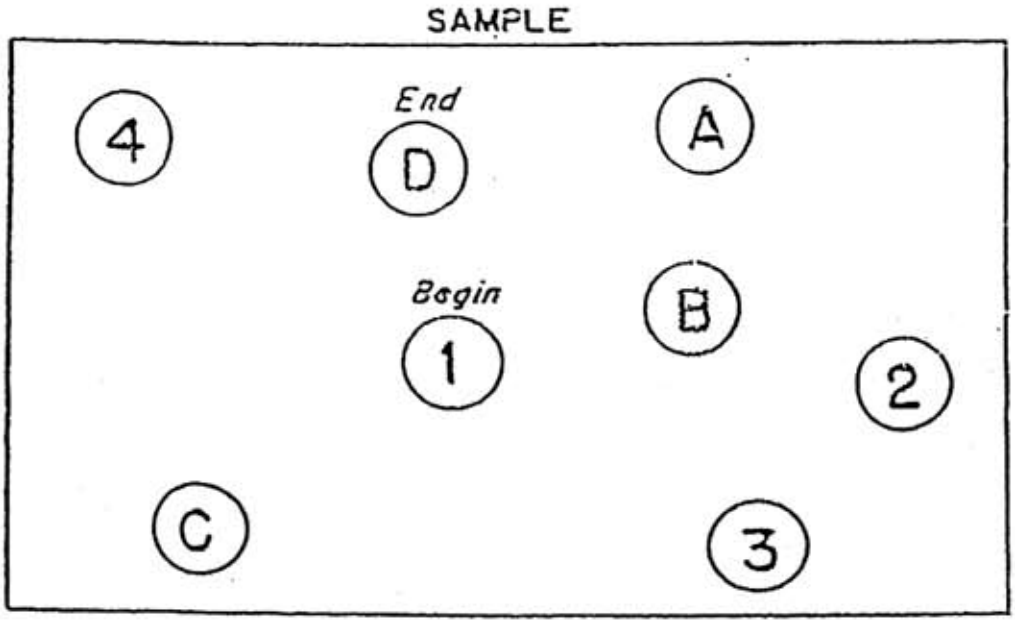
SAMPLE

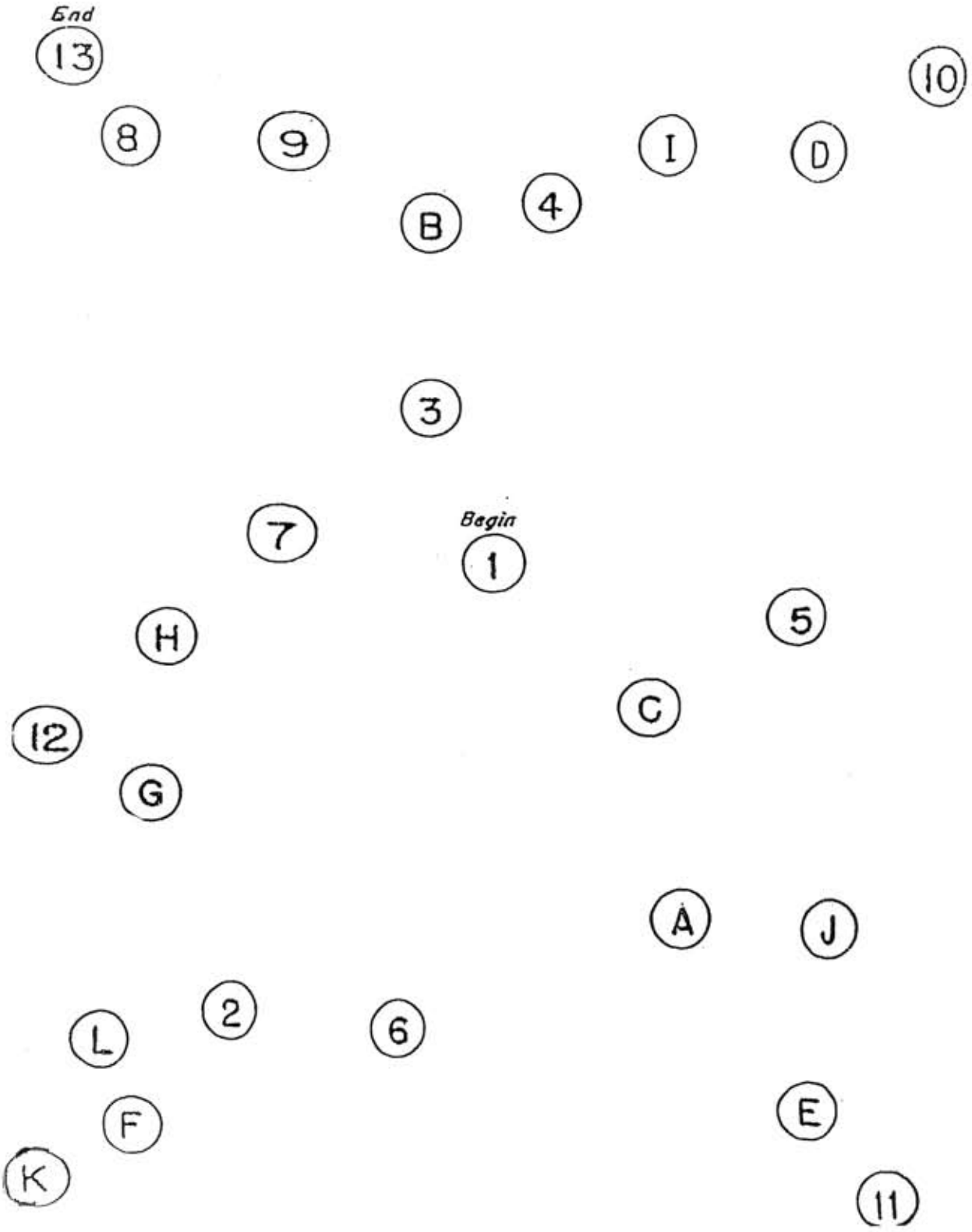



17
21
15
20
19
16
18
4
22
5
13
6
Begin
1
24
7
14
2
8
10
3
9
End
25
11
12
23

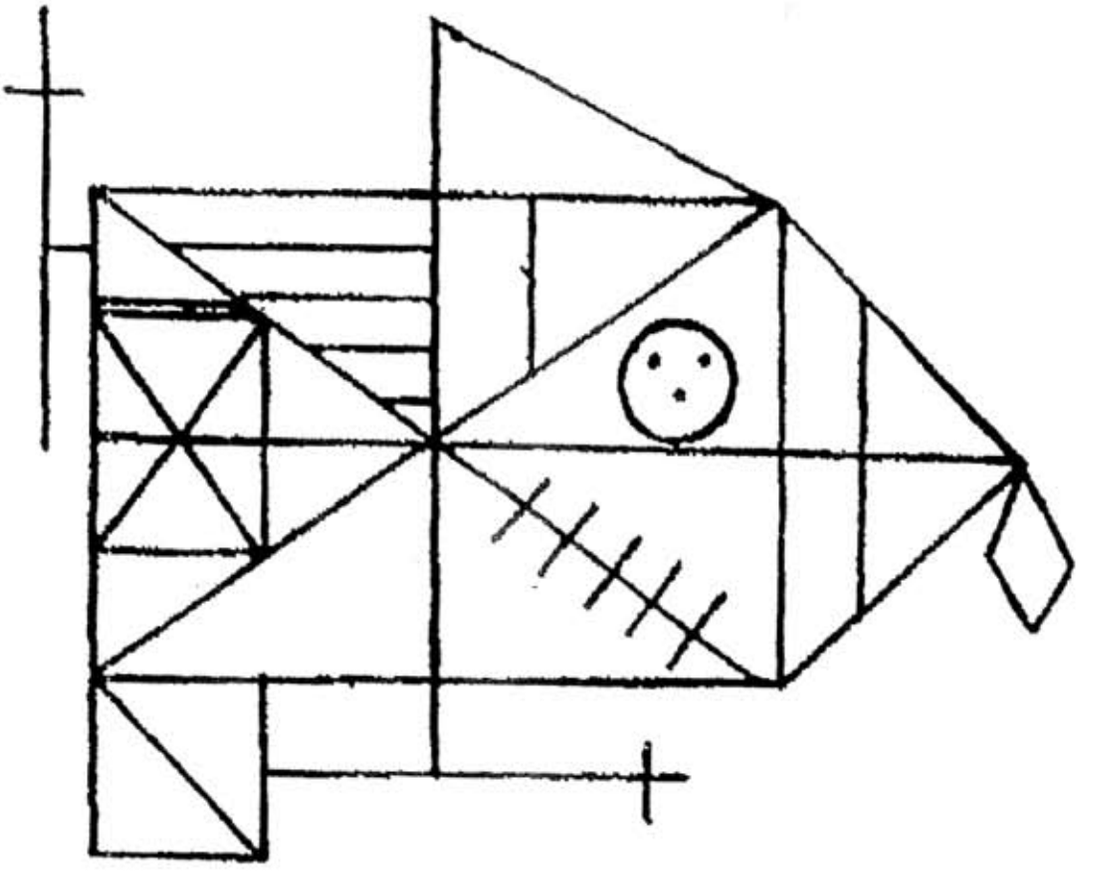
TRAIL MAKING

Part B

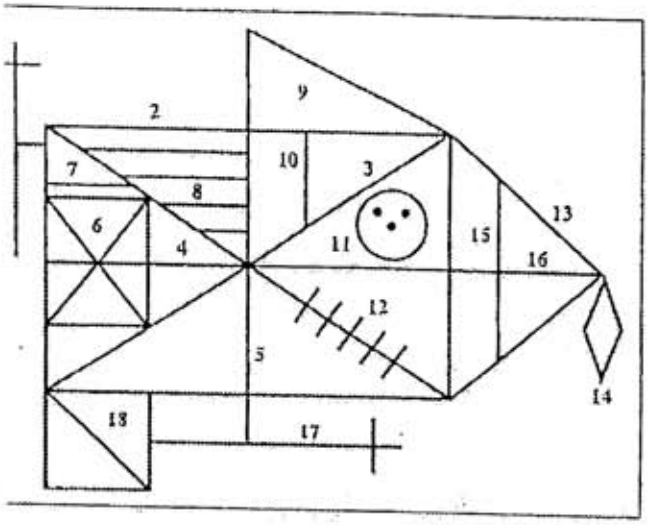



End
13
10
8
9
I
D
4
B
3
7
Begin
1
5
H
C
12
G
A
J
2
6
L
E
F
K
11

Rey Complex Figure Task



Scoring Sheet



Scoring Criteria for RCFT Drawings

Score	Accuracy	Placement
2	Accurately drawn	Correctly placed
1	Accurately drawn	Incorrectly placed
1	Inaccurately drawn	Correctly placed
0.5	Inaccurately drawn, but recognizable	Incorrectly placed
0	Inaccurately drawn and unrecognizable, or omitted	Incorrectly placed

Scoring Element	Copy	Immediate Recall	Delayed Recall
1 Vertical Cross	2 1 0.5 0	2 1 0.5 0	2 1 0.5 0
2 Large Rectangle	2 1 0.5 0	2 1 0.5 0	2 1 0.5 0
3 Diagonal Cross	2 1 0.5 0	2 1 0.5 0	2 1 0.5 0
4 Horizontal Midline of Large Rectangle (2)	2 1 0.5 0	2 1 0.5 0	2 1 0.5 0
5 Vertical Midline of Large Rectangle (2)	2 1 0.5 0	2 1 0.5 0	2 1 0.5 0
6 Small Rectangle	2 1 0.5 0	2 1 0.5 0	2 1 0.5 0
7 Small Horizontal Line above Small Rectangle (6)	2 1 0.5 0	2 1 0.5 0	2 1 0.5 0
8 Four Parallel Lines	2 1 0.5 0	2 1 0.5 0	2 1 0.5 0
9 Small Triangle above Large Rectangle (2)	2 1 0.5 0	2 1 0.5 0	2 1 0.5 0
10 Small Vertical Line within Large Rectangle (2)	2 1 0.5 0	2 1 0.5 0	2 1 0.5 0
11 Circle with Three Dots	2 1 0.5 0	2 1 0.5 0	2 1 0.5 0
12 Five Parallel Lines	2 1 0.5 0	2 1 0.5 0	2 1 0.5 0
13 Sides of Large Triangle attached to Large Rectangle (2)	2 1 0.5 0	2 1 0.5 0	2 1 0.5 0
14 Diamond	2 1 0.5 0	2 1 0.5 0	2 1 0.5 0
15 Vertical Line within Sides of Large Triangle (13)	2 1 0.5 0	2 1 0.5 0	2 1 0.5 0
16 Horizontal Line within Sides of Large Triangle (13)	2 1 0.5 0	2 1 0.5 0	2 1 0.5 0
17 Horizontal Cross	2 1 0.5 0	2 1 0.5 0	2 1 0.5 0
18 Square attached to Large Rectangle (2)	2 1 0.5 0	2 1 0.5 0	2 1 0.5 0
	_____	_____	_____
	Raw score	Raw score	Raw score

Recognition Trial Worksheet

_____ Recognition True Positives = Sum of items 2, 5, 7, 8, 9, 12, 13, 15, 19, 20, 22, and 24 that were circled.

_____ Recognition False Positives = Sum of items 1, 3, 4, 6, 10, 11, 14, 16, 17, 18, 21, and 23 that were circled.

_____ Recognition True Negatives = 12 minus Recognition False Positives.

_____ Recognition False Negatives = 12 minus Recognition True Positives.

_____ Recognition Total Correct = Recognition True Positives plus Recognition True Negatives.

Measurement of Everyday Functioning (Farias et al., 2006)

Please rate the following statements on the scale provided by comparing to abilities 10 years ago

Everyday Functioning Items	Doesn't Apply	Better or no change	Questionable or occasional problems	Consistently a little worse	Consistently much worse
Everyday Memory					
1. Remembering a few shopping items without a list	0	1	2	3	4
2. Remembering the names of new people	0	1	2	3	4
3. Remembering things that happened recently (such as recent outings, events in the news)	0	1	2	3	4
4. Recalling conversations a few days earlier	0	1	2	3	4
5. Remembering where he/she placed objects	0	1	2	3	4
6. Remembering movies or TV shows he/she has seen before	0	1	2	3	4
7. Repeating stories and/or questions	0	1	2	3	4
8. Remembering the current date or day of the week	0	1	2	3	4
9. Remembering he/she has already told someone something	0	1	2	3	4
10. Remembering thins about family and friends (such as addresses, birthdates, jobs)	0	1	2	3	4
11. Remembering appointments, meetings or engagements	0	1	2	3	4
Everyday Language					
12. Following a story in a book	0	1	2	3	4
13. Describing a program he/she has watched on TV	0	1	2	3	4
14. Understanding spoken directions or instructions	0	1	2	3	4
15. Understand what he/she has read	0	1	2	3	4
Everyday Planning					
16. Planning a big dinner, social event, birthday party, or	0	1	2	3	4

club meeting					
17. The ability to develop a good strategy in a game of skill (ie. Card or board game)	0	1	2	3	4
18. Planning a weekly meal schedule ahead of time	0	1	2	3	4
19. Planning the sequence of stops on a shopping trip	0	1	2	3	4
20. Planning recreational outings such as family visits, golfing, picnics, weekend getaways)	0	1	2	3	4
21. Developing a schedule in advance of anticipated events	0	1	2	3	4
22. Thinking ahead	0	1	2	3	4
23. Thinking things through before acting	0	1	2	3	4
Everyday Organisation					
24. Assembling business, tax or financial records	0	1	2	3	4
25. Balancing the chequebook without error	0	1	2	3	4
26. Keeping financial records organised	0	1	2	3	4
27. Prioritising tasks by importance	0	1	2	3	4
28. Keeping mail and papers organised	0	1	2	3	4
Everyday Divided Attention					
29. The ability to do 2 things at once	0	1	2	3	4
30. Returning to a task after being interrupted	0	1	2	3	4

Instrumental Activities of Daily Living Scale (IADL)	Rater's Initials	Date
		Rating
A. ABILITY TO USE THE TELEPHONE 0=Not applicable 1=Operates telephone on own initiative 2=Dials a few well know numbers 3=Answers telephone but does not dial 4=Does not use telephone at all		
B. SHOPPING 0=Not applicable. Patient did not perform this task previously. 1=Takes care of all shopping needs independently 2=Shops independently for small purchases 3=Needs to be accompanied on a shopping trip 4=Completely unable to shop.		
C. FOOD PRERARATION 0=Not applicable. Patient did not perform this task previously. 1=Plans, prepares and serves adequate meals independently. 2=Prepares adequate meals if supplied with ingredients 3=Heats and serves prepared meals but does not maintain an adequate diet 4=Needs to have meals served and prepares		
D. HOUSEKEEPING 0=Not applicable. Patient did not perform this task previously 1=Maintains house alone or with occasional assistance (eg. Heavy work – domestic help) 2=Performs light daily tasks such as dish washing and bed making. 3=Performs light daily tasks but cannot maintain acceptable levels of cleanliness 4=Needs help with all home maintenance tasks 5=Does not participate in any housekeeping tasks		
E. LAUNDRY 0=Not applicable. Patient did not perform this task previously 1=Does personal laundry completely 2=Launders small items – rinses socks, stockings, etc. 3=All laundry must be done by other		
F. MODE OF TRANSPORTATION 0=Not applicable. Patient did not perform this task previously 1=Travels independently on public transportation or drives own car. 2=Arranges own travel via taxi, but does not otherwise use public transportation 3=Travels on public transportation when assisted or accompanied by another 4=Travel limited to taxi or automobile with the assistance of another 5=Does not travel at all		
G. RESPONSIBILITY FOR OWN MEDICATIONS 0=Not applicable. Patient did not perform this task previously 1=Is responsible for taking medication in correct dosages at the correct time 2=Takes responsibility if medication is prepared in advance in separate dosages 3=No longer dispenses own medication		
H. ABILITY TO HANDLE FINANCES 0=Not applicable. Patient did not perform this task previously 1=Manages financial matters independently (budgets, writes cheques, pays rent, goes to bank) collects and keeps track of income 2=Manages day-to-day purchases but needs help with banking, major purchases, etc. 3=No longer handles money		

Personal & Self Maintenence Scale

PSMS	Rater's Initials	Date
		Rating
A. TOILET 1=Cares for self at toilet completely, no incontinence. 2=Needs to be reminded or needs help in cleaning self, or has rare (weekly at most) accidents 3=Soiling or wetting while asleep more than once a week 4=Soiling or wetting while awake more than once a week 5=No control of bowels or bladder		
B. FEEDING 1=Eats without assistance 2=Eats with minor assistance at meal times and/or with special preparation of food or help in cleaning up after meals. 3=Feeds self with moderate assistance or is untidy 4=Requires extensive assistance for meals 5=Does not feed self and resists all efforts of others to feed him/her.		
C. DRESSING 1=Dresses, undresses and selects clothing from own wardrobe 2=Dresses and undresses self, with minor assistance 3=Needs moderate assistance in dressing or selection of clothes 4=Needs major assistance in dressing but co-operative with the efforts of others to help 5=Completely unable to dress self and resists efforts of others to help		
D. GROOMING (neatness, hair, hands, face, clothing) 1=Always neatly dressed, well-groomed, without assistance 2=Grooms self adequately with occasional minor assistance, e.g. shaving 3=Needs moderate and regular assistance or supervision in grooming. 4=Needs total grooming care, but can remain well-groomed after help from others. 5=Actively negates all efforts of others to maintain grooming.		
E. PHYSICAL AMBULATION 1=Goes about grounds or city 2=Ambulates within residence or about one block distant 3=Ambulates with assistance of (Check one): a. Another person b. Railing c. Cane d. Walker e. Wheelchair (check one) 1. Gets in and out without help 2. Needs help getting in, out. 4=Sits unsupported in chair or wheelchair, but cannot propel self without help 5=Bedridden more than half the time.		
F. BATHING 1=Bathes self (tub, shower, sponge-bath) without help 2=Bathes self with help in getting in and out of tub 3=Washes face and hands only, but cannot bathe the rest of the body 4=Does not self wash, but is co-operative with those who bathe him/her 5=Does not try to wash self and resists efforts to keep him/her clean		

GERIATRIC DEPRESSION SCALE

DIRECTIONS: Say, **"Choose the best answer for how you felt over the <u>past week</u>".**

1. Are you basically satisfied with your life? ..... Yes/**No**
2. Have you dropped many of your activities and interests? ..... **Yes**/No
3. Do you feel that your life is empty? ..... **Yes**/No
4. Do you often get bored? ..... **Yes**/No
5. Are you hopeful about the future? ..... Yes/**No**
6. Are you bothered by thoughts you can't get out of your head? ..... **Yes**/No
7. Are you in good spirits most of the time? ..... Yes/**No**
8. Are you afraid that something bad is going to happen to you? ..... **Yes**/No
9. Do you feel happy most of the time? ..... Yes/**No**
10. Do you often feel helpless? ..... **Yes**/No
11. Do you often get restless and fidgety? ..... **Yes**/No
12. Do you prefer to stay home rather than go out and do new things? ..... **Yes**/No
13. Do you frequently worry about the future? ..... **Yes**/No
14. Do you feel that you have more problems with memory than most? ..... **Yes**/No
15. Do you think it is wonderful to be alive now? ..... Yes/**No**
16. Do you often feel down hearted and blue? ..... **Yes**/No
17. Do you feel pretty worthless the way you are now? ..... **Yes**/No
18. Do you worry a lot about the past? ..... **Yes**/No
19. Do you find life very exciting? ..... Yes/**No**
20. Is it hard to get started on new projects? ..... **Yes**/No
21. Do you feel full of energy? ..... Yes/**No**
22. Do you feel that your situation is hopeless? ..... **Yes**/No
23. Do you think that most people are better off than you are? ..... **Yes**/No
24. Do you frequently get upset over little things? ..... **Yes**/No
25. Do you frequently feel like crying? ..... **Yes**/No
26. Do you have trouble concentrating? ..... **Yes**/No
27. Do you enjoy getting up in the morning? ..... Yes/**No**
28. Do you prefer to avoid social gatherings? ..... **Yes**/No
29. Is it easy for you to make decisions? ..... Yes/**No**
30. Is your mind as clear as it used to be? ..... Yes/**No**

Depression Score: ☐☐/30